Functional reprogramming of *Candida glabrata* epithelial adhesins by exchange of variable structural motifs

---

Dissertation

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer. nat.)

des Fachbereichs Biologie der Philipps-Universität Marburg

vorgelegt von

Daniel Hoffmann

aus

Warburg (Westfalen)

Marburg/Lahn 2020

Die vorliegende Arbeit wurde in der Zeit von März 2016 bis August 2020 am Fachbereich Biologie unter Leitung von Herrn Prof. Dr. Hans-Ulrich Mösch angefertigt.

Vom Fachbereich Biologie der Philipps-Universität Marburg
(Hochschulkennziffer 1180) als Dissertation angenommen am: _____

Erstgutatchter: Prof. Dr. Hans-Ulrich Mösch
Zweitgutatchter: Prof. Dr. Lars-Oliver Essen

Tag der Disputation: _____

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit mit dem Titel “Functional reprogramming of *Candida glabrata* epithelial adhesins by exchange of variable structural motifs” selbstständig und ohne unerlaubte Hilfe angefertigt habe. Außerdem habe ich mich keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient.

Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Ort, Datum	Daniel Hoffmann

Publications

Brückner, S., Schubert, R., Kraushaar, T., Hartmann, R., Hoffmann, D., Jelli, E., Drescher, K., Müller, D. J., Essen, L.-O., and Mösch, H.-U. (2020). Kin Discrimination in Social Yeast Is Mediated by Cell Surface Receptors of the Flo11 Adhesin Family. *eLife* **9**.

Hoffmann, D., Diderrich, R., Reithofer, V., Friederichs, S., Kock, M., Essen, L.-O., and Mösch, H.-U. (2020). Functional Reprogramming of *Candida glabrata* Epithelial Adhesins: The Role of Conserved and Variable Structural Motifs in Ligand Binding. *J. Biol. Chem.* **295** (35):12512–12524.

Poster

Hoffmann, D., Diderrich, R., Friederichs, S., Kock, M., Essen, L.-O., Mösch, H.-U. (2019). Reprogramming of epithelial adhesin ligand binding specificity by exchange of variable structural motifs. *International Conference on Yeast Genetics and Molecular Biology*. Gothenburg, Sweden

Crystal structures

The atomic coordinates and structure factors for Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} have been deposited in the Protein Data Bank (www.wwpdb.org) under the following accession codes:

Hoffmann, D., Diderrich, R., Kock, M., Friederichs, S., Reithofer, V., Essen, L.-O. and Moesch, H.-U. (2020). *J. Biol. Chem.*

6Y9J Crystal Structure of subtype-switched Epithelial Adhesin 1 to 9 A domain (Epa1-CBL2Epa9) from *Candida glabrata* in complex with beta-lactose

6Y98 Crystal Structure of subtype-switched Epithelial Adhesin 9 to 1 A domain (Epa9-CBL2Epa1) from *Candida glabrata* in complex with beta-lactose

"You may test that assumption
at your convenience."

Jean-Luc Picard

Summary

The yeast *Candida glabrata* is part of the human microbiome and usually employs a commensal lifestyle, but this fungus is also able to act as an opportunistic pathogen, causing localized as well as severe systemic infections. For host invasion and dissemination, *C. glabrata* disposes of a large number of cell wall attached proteins, the most prominent of which are the epithelial adhesins (Epas). The Epa family encompasses more than 20 members, which act as lectins. All Epa paralogs share the common tripartite architecture of fungal adhesins, composed of an N-terminal A domain (adhesion domain), a central B domain consisting of a variable number of serine- and threonine-rich repeats, and a C-terminal region carrying a glycosylphosphatidylinositol (GPI) anchor for attachment to the cell wall. The lectin function of Epa adhesins is conferred by a combination of conserved and variable structural elements within their A domains. Together, these elements form an inner and outer binding pocket and are thought to control ligand binding affinity and specificity.

In this work, variable structural elements of several Epa paralogs were functionally characterized using structure-based mutational analysis, to precisely elucidate their role in conferring host cell adhesion and ligand binding specificity. For this purpose, an array of chimeric EpaA variants carrying directed exchanges of highly variable regions in the inner and the outer binding pocket were constructed and functionally characterized. I*n vivo* adhesion assays with human epithelial cells revealed that both of these structural elements are involved in host cell binding. Specifically, exchanges within the inner binding pocket resulted in a lower binding strength. In contrast, the exchange of elongated loops in the outer binding pocket for shorter variants showed a significant increase in host cell adhesion, whereas chimeras carrying longer instead of shorter loops exhibited reduced adhesion. Chimeric EpaA variants were further characterized by glycan array analysis and fluorescence titration spectroscopy. These measurements demonstrated that the ligand binding specificity of EpaA domains can in principal be reprogrammed by exchange of structural elements in the inner binding pocket, with albeit limited predictability. In contrast, exchanges of outer binding pocket elements generally did not affect ligand binding patterns. For further structural characterization of elongated loops in the outer binding pocket, soaking experiments were performed using protein crystals and complex glycan structures. Since this approach did not yield structural data, the flexibility of long loops was analyzed by molecular dynamic simulations, in order to test a putative lid functionality. These simulations showed that in the absence of glycan ligands, the elongated loop can principally adopt a stable conformation, but does not cover the binding pocket. In the presence of a tetrameric glycan, however, the reducing end of the ligand was stabilized by direct contact with the loop, indicating a crucial function of this variable structural element in binding complex glycan structures.

In a further part of this thesis, the function of variable amino acid residues within the inner binding pocket was investigated which have been postulated to confer specific binding of sulfated glycans. To test this hypothesis, corresponding residues were functionally characterized by

mutational analysis in combination with *in vivo* adhesion tests to human epithelial cells and *in vitro* ligand binding studies. Interestingly, no correlation was detected between mutated positions and specific sulfoglycan binding. However, docking simulations with sulfated disaccharide ligands suggest that other steric effects control the precise fitting of spatially demanding sulfate groups into the binding pockets of EpaA domains.

In summary, results obtained in this work support the view, that variation of several structural elements in the inner and outer ligand binding pocket of Epa adhesins is a main driver of their functional diversification and evolution.

Zusammenfassung

Der Hefepilz *Candida glabrata* ist Teil des menschlichen Mikrobioms und zeigt üblicherweise eine kommensale Lebensweise. Dieser Pilz ist aber auch in der Lage, als opportunistischer Erreger zu wirken und lokale sowie schwere systemische Infektionen zu verursachen. Für das Eindringen in den Wirt und die weitere Ausbreitung verfügt *C. glabrata* über eine große Anzahl zellwandgebundener Proteine, von denen die epithelialen Adhäsine (Epas) die bekanntesten sind. Die Epa-Familie umfasst mehr als 20 Mitglieder, die als Lektine wirken. Alle Epa-Paraloge besitzen die dreiteilige Architektur pilzlicher Adhäsine mit einer N-terminalen A-Domäne (Adhäsionsdomäne) und einer zentralen B-Domäne, die aus einer variablen Anzahl von serin- und threonin-reichen Wiederholungen besteht. Die C-terminale Region trägt einen Glycosylphosphatidylinositol (GPI)-Anker zur Befestigung an der Zellwand. Die Lektinfunktion von Epa-Adhäsinen wird durch eine Kombination von konservierten und variablen Strukturelementen innerhalb ihrer A-Domäne bestimmt. Zusammen bilden diese Elemente eine innere und äußere Bindetasche und kontrollieren voraussichtlich die Affinität und Spezifität der Ligandenbindung.

In dieser Arbeit wurden variable Strukturelemente mehrerer Epa-Paraloge mit Hilfe einer strukturbasierten Mutationsanalyse funktionell charakterisiert, um ihre Rolle bei der Vermittlung der Wirtszelladhäsion und der Spezifität der Ligandenbindung aufzuklären. Zu diesem Zweck wurde eine Reihe chimärer EpaA-Varianten konstruiert und funktionell charakterisiert, in denen hochvariable Regionen der inneren und äußeren Bindetasche ausgetauscht wurden. *In-vivo*-Adhäsionsassays mit menschlichen Epithelzellen zeigten, dass diese beiden Strukturelemente an der Wirtszellbindung beteiligt sind. Insbesondere führten Austausche innerhalb der inneren Bindetasche zu einer verringerten Bindungsstärke. Im Gegensatz dazu zeigte der Austausch verlängerter Schleifen in der äußeren Bindetasche gegen kürzere Varianten eine signifikante Zunahme der Wirtszelladhäsion, während Chimären, die längere statt einer kürzeren Schleife trugen, eine geringere Adhäsion aufwiesen. Die chimären EpaA-Varianten wurden anschließend durch Glykanarray-Analyse und Fluoreszenztitrationsspektroskopie charakterisiert. Diese Messungen zeigten, dass die Ligandenbindungsspezifität der EpaA-Domänen prinzipiell durch Austausch von Strukturelementen in der inneren Bindetasche umprogrammiert werden kann, allerdings mit begrenzter Vorhersagbarkeit. Im Gegensatz dazu hatte der Austausch von Elementen in der äußeren Bindetasche im Allgemeinen keinen Einfluss auf die Ligandenbindungsmuster. Zur weiteren strukturellen Charakterisierung der verlängerten Schleifen in der äußeren Bindetasche wurden *Soaking*-Versuche mit Proteinkristallen und komplexen Glykanen durchgeführt. Da dieser Ansatz keine strukturellen Daten lieferte, wurde die Flexibilität der verlängerten Schleifen durch Molekulardynamiksimulationen analysiert, um eine vermutete Deckelfunktionalität zu testen. Diese Simulationen zeigen, dass in Abwesenheit von Glykanliganden die verlängerte Schleife prinzipiell eine stabile Konformation annehmen kann, aber die Bindetasche nicht bedeckt. In Gegenwart eines tetrameren Glykans wurde jedoch das reduzierende Ende des Liganden durch direkten Kontakt mit der Schleife stabilisiert, was auf eine wichtige Funktion dieses variablen

Strukturelements bei der Bindung komplexer Glykanstrukturen hinweist.

In einem weiteren Teil dieser Arbeit wurde die Funktion variabler Aminosäurereste innerhalb der inneren Bindetasche untersucht, für die postuliert wurde, dass sie für eine spezifische Bindung von sulfatierten Glykanen verantwortlich sind. Um diese Hypothese zu überprüfen, wurden entsprechende Auminosäuren durch Mutationsanalyse in Kombination mit *in-vivo*-Adhäsionstests an humanen Epithelzellen und in-vitro-Ligandenbindungsstudien funktionell charakterisiert. Interessanterweise wurde keine Korrelation zwischen den mutierten Positionen und spezifischer Sulfoglykanbindung festgestellt. Bindungssimulationen mit sulfatierten Disaccharidliganden deuten jedoch darauf hin, dass andere sterische Effekte die spezifische Bindung von räumlich anspruchsvollen Sulfatgruppen in EpaA-Domänen kontrollieren.

Zusammenfassend unterstützen die in dieser Arbeit erzielten Ergebnisse die Annahme, dass die Variation verschiedener Strukturelemente in der inneren und äußeren Ligandenbindetasche der EpaA-Adhäsine ein Hauptfaktor für ihre funktionelle Diversifizierung und Evolution ist.

Table of Contents

1 Introduction

1.1 Yeasts are a diverse group of fungi

Yeasts are unicellular eukaryotic organisms that belong to the kingdom of fungi and can be found in almost all ecosystems on earth. They are able to colonize different terrestrial and aquatic environments ranging from cold or dry habitats to the deep sea (Grossart et al. 2019). The multitude of habitats is accompanied by a diverse set of lifestyles which include fermentation of floral nectar, the formation of symbiotic relationships with plants and animals or even predatory behavior (Nguyen et al. 2007; Liti 2015; Junker et al. 2018). Altogether, about 1500 yeast species are known that are placed in the phyla of ascomycetes and basidiomycetes (Kurtzman and Piškur 2006). Yeasts commonly utilize mono- or disaccharide sugars as their main carbon source (Flores et al. 2000) and often are capable of performing fermentation in anaerobic or even aerobic conditions (Merico et al. 2007). For reproduction, yeasts generally multiply by vegetative growth, which occurs asexually either by budding or fission, with budding being the most common form of asexual reproduction (Balasubramanian et al. 2004). Many species are also capable of sexual reproduction which provides advantages for adaptation by genetic recombination (Kleiman and Tannenbaum 2009). Furthermore, the ability of diploid yeast cells to create haploid spores during meiosis extends their resistance against external stresses in comparison to haploid vegetative cells (Briza et al. 1990).

Several yeast species are important for humans as they are used for the production of food, medication or fuel and are also used as eukaryotic model organisms in basic research. One of the most widely used species is the baker's yeast *Saccharomyces cerevisiae*, which has been used for at least 7000 to 8000 years for the production of fermented foods and alcoholic beverages (Aouizerat et al. 2019). It is also used as a model organism for scientific research and was the first eukaryotic organism whose genome has been fully sequenced (Goffeau et al. 1996) which enabled a targeted investigation of protein functions and signal transduction pathways (Winzeler et al. 1999; Giaever et al. 2002). Other species like *Yarrowia lipolytica* recently have emerged as bioreactors for industrial production of biofuel, biodegradable polymers or polyketides (Madzak 2018).

In addition to their economic value, a number of different yeast species are part of the human microbiome and therefore also relevant for health-related topics. Normally, these organisms are commensals that colonize different parts of the human body like the skin or mucosal surfaces in the oral cavity or gut (Tam et al. 2015), but they can also act as opportunistic pathogens. Under certain circumstances, such as in patients with a weakened immune system, these pathogens can cause superficial and invasive infections that can be hard to treat and thus present a serious thread to their host (Sims et al. 2005). Parameters that complicate the treatment of fungal infections are for example the expression of virulence factors, formation of biofilms or natural resistances against certain antimycotica (Ganguly and Mitchell 2011). The mortality of systemic fungal

infections is up to 50 % with an estimated number of 1.5 million deaths per year (Brown et al. 2012), which makes them a substantial problem in affected patients. Among the most commonly found fungi associated with infections in human hosts are different *Candida* species with *Candida albicans* and *Candida glabrata* being the two most abundant representatives (Pfaller et al. 2011; Diekema et al. 2012; Guinea 2014).

1.2 *Candida glabrata* – commensal and pathogen

The genus *Candida* comprises a number of different yeasts, several of which have been shown to be commensal organisms of the human body but also be able to act as opportunistic human pathogens. Phylogenetic analyses have revealed, that this diverse group of yeasts separates into two different clades (Gabaldón et al. 2013). The *Nakaseomyces* clade comprises *C. glabrata* and other closely related species and has emerged after a genome duplication that has happened in a common ancestor of *C. glabrata* and *S. cerevisiae* (Figure 1). Other *Candida* species, including the relevant human pathogens *C. albicans* and *C. tropicalis*, show a specific codon usage that translates the CTG codon into leucine instead of serine which places these species into the so called CTG clade (Massey et al. 2003).

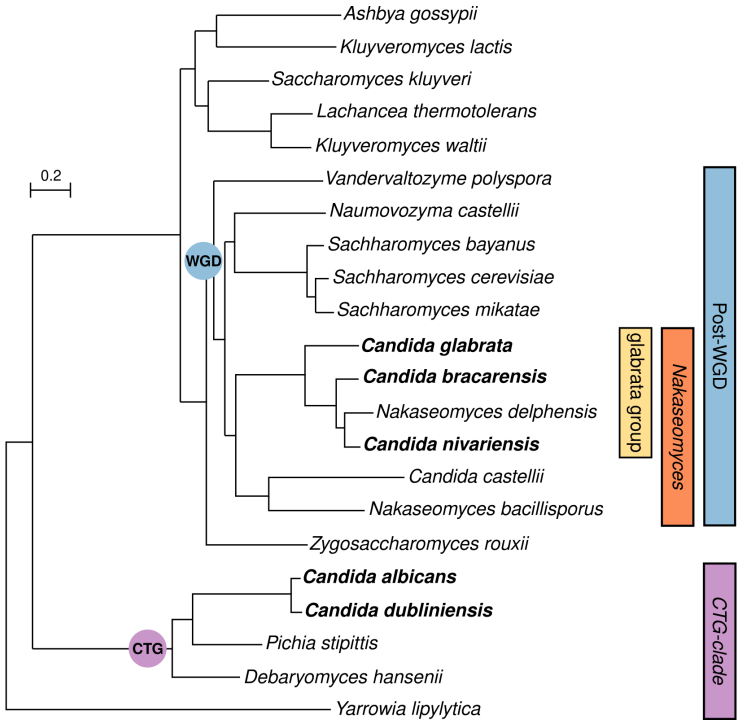

Figure 1: Phylogenetic tree showing the relationship of different yeast species from the Saccharomycotina subphylum.
Phylogenetic analysis of *C. glabrata* and other Saccharomycotina species showed that *C. glabrata* and *C. albicans* belong to different clades despite being assigned to the same genus. The CTG clade (purple) comprises yeasts that translate the CTG codon to leucine instead of serine while the common ancestor of the WGD clade (blue) experienced a whole genome duplication. Bold names indicate pathogenic species. Their assignment to different clades shows that pathogenicity is a paraphyletic trait. The figure was adapted from Gabaldón and Carreté 2016.

Differing and common features between species of both clades can be emphasized best by a direct comparison between *C. glabrata* and *C. albicans*. One major difference is ploidy and overall genome size with *C. glabrata* having a haploid genome of 12.3 Mb while *C. albicans* is a diploid organism with a genome size of 15.4 Mb (Kaur et al. 2005). This is also reflected by their cell size and morphology with *C. albicans* being significantly larger (4 to 6 µm) than *C. glabrata* (1 to 4 µm) and also being able to form true hyphae and pseudohyphae, while *C. glabrata* mainly grows in its yeast form and only shows pseudohyphal growth under certain low nitrogen conditions (Fidel et al. 1999; Csank and Haynes 2000). Both organisms have two mating types and carry a full set of genes necessary for maintaining a sexual cycle (Fabre et al. 2005). However, in contrast to *C. albicans* the sexual cycle of *C. glabrata* remains cryptic as evidence for mating in *C. glabrata* has not yet been found. In fact, it has been shown that artificial induction of mating type switching can lead to cell death in *C. glabrata* (Boisnard et al. 2015). Two features that further distinguish both species are the innate resistance of *C. glabrata* against azole-based antimycotica (Castanheira et al. 2014) and its auxotrophy for niacin, thiamine and pyridoxine which is compensated by uptake from the human host (Rodrigues et al. 2014). In contrast, *C. albicans* is prototrophic and also does not have any innate resistance against antifungal drugs.

While both species show a number of differences, they also share certain features like the adaption to a growth temperature of 37 °C, phenotypic switching (Brockert et al. 2003), the formation of biofilms (Cuéllar-Cruz et al. 2012) and their lifestyle as an opportunistic pathogen. Moreover, in their human host both species mainly infect mucosal surfaces of the oral cavity, the gastrointestinal tract and the vagina (Kaur et al. 2005). The formation of biofilms on mucosal surfaces makes *Candida* infections hard to treat, as their extracellular matrix prevents antifungal drugs and immune cells from reaching the yeast cells (Figure 2). Biofilms also present serious problems for clinical environments where both, *C. albicans* and *C. glabrata*, are able to colonize abiotic surfaces like catheters, implants or dentures (Douglas 2003). This renders them a life threatening risk for immunocompromised patients because these biofilms can disseminate and cause systemic infections with possibly fatal consequences for affected hosts (Hickey et al. 1983).

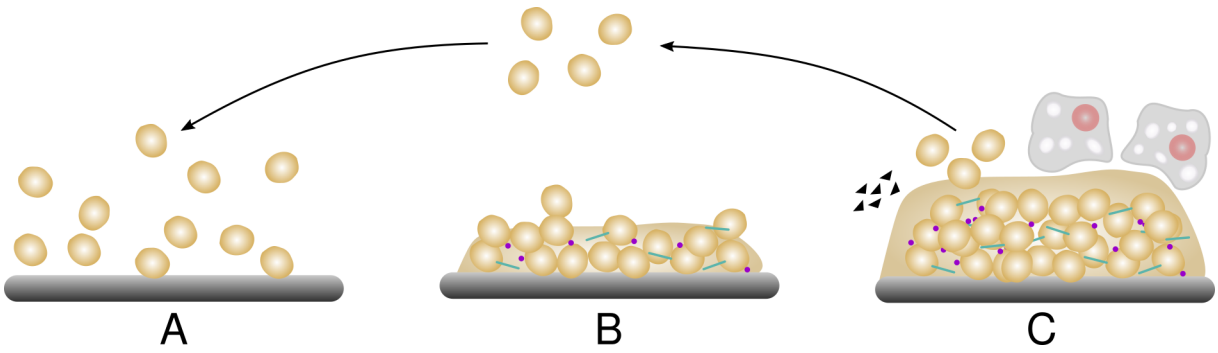

Figure 2: Biofilm formation in *C. glabrata*.
A. Yeast cells start to adhere to biotic or abiotic substrates. *B.* Adherent cells start to proliferate and secrete polysaccharides, carbohydrates, proteins and other components that form an extracellular matrix. *C* Inside the matured biofilm *C. glabrata* is protected against host immune cells and antifungal drugs (black triangles). At this stage of biofilm formation, single cells disseminate and colonize new infection sites.

About 65 % of all worldwide *Candida* infections are caused by *C. albicans* and *C. glabrata* with *C. glabrata* infections showing a slow but steady increase in numbers (Pfaller et al. 2019). Interestingly, both species do not necessarily compete for the colonization of their host despite being found at the same body sites. In fact, mixed infections with both pathogens can often be found (Vazquez 1999; Redding et al. 2004) which actually are more severe than infections that involve only a single *Candida* species (Redding et al. 2000; Coco et al. 2008). For oral infections it has been shown that *C. glabrata* cannot invade the host epithelium, but has the ability to bind to *C. albicans* hyphae (Figure 3) and thus passively enter the host tissue (Tati et al. 2016).

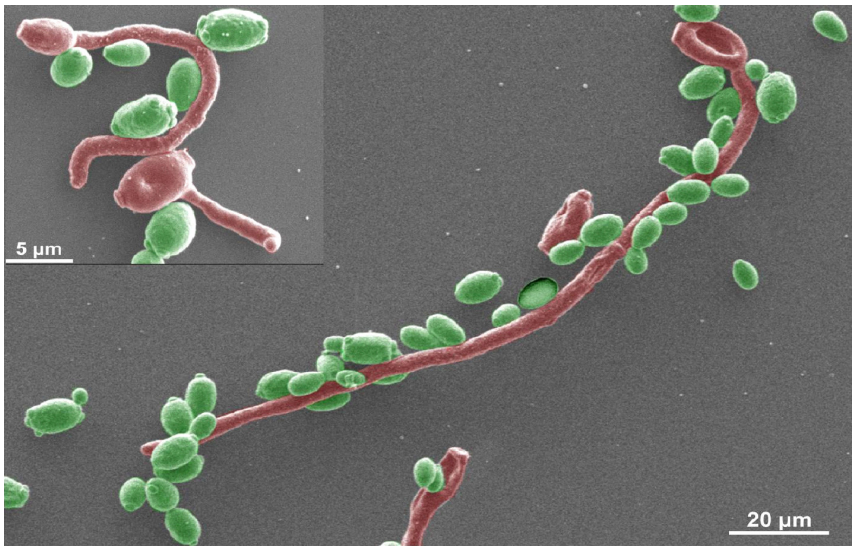

Figure 3: *C. glabrata* adhering to *C. albicans* hyphae.
Transmission electron image of *C. glabrata* cells (green) adhering to *C. albicans* hyphae (red). By adherence along the length of the hyphae, *C. glabrata* is able to penetrate host tissue and cause disseminated infections. The figure was adapted from Tati et al. 2016.

This underlines the role of *C. glabrata* as a commensal organism, which occasionally also acts as a highly opportunistic pathogen. Another specialization of *C. glabrata* is its ability to persist and even proliferate inside macrophages (Kasper et al. 2015). This presents a different approach to other yeast pathogens that usually try to evade the host immune system or escape from phagosomes by killing the macrophage (Uwamahoro et al. 2014). Usually, phagocytized pathogens are quickly killed by acidification and maturation of the phagosome. However, *C. glabrata* can prevent the maturation of phagosomes and even multiply inside the compartment until the macrophage lyses due to fungal load (Kasper et al. 2015). This effectively hides the yeast cells and enables them to evade further host immune reactions. To date is is not fully understood how *C. glabrata* can, despite its lack of hyphal growth, invade human tissue. It might be possible that *C. glabrata* uses the ability to survive inside macrophages to cross epithelial barriers, a strategy that has already been found in *Cryptococcus neoformans*, another fungal pathogen (Charlier et al. 2009).

Despite its ability to colonize hosts by taking advantage of other pathogens or immune cells, *C. glabrata* is only able to effectively infect hosts that are immunocompromised. In healthy individuals, nutrient depletion and secretion of chemical factors by the host microbiome

effectively limits growth of fungal pathogens and thus contribute to the prevention of infections (Hall and Noverr 2017). Therefore, extensive use of antibiotics is one of the most common risk factors for fungemia (Sims et al. 2005) as it allows *C. glabrata* to proliferate and attach to a larger number of host cells. As the adhesion to its host is a crucial factor for further colonization, the cell wall of *C. glabrata* has a central role for a number of steps that define its pathogenicity.

1.3 Composition and function of the yeast cell wall

The yeast cell wall is a structurally complex cellular component that has crucial functions for fungal cells. It provides a robust shell that withstands the turgor pressure of the cell and determines its shape. Furthermore, it serves as a protective barrier against harmful conditions like heat, cold, desiccation, and osmotic stress (Free 2013). Changes in the cellular environment can be detected by sensor proteins which allows the cell to react accordingly. In *Candida* species the cell wall consists of two layers (Figure 4) with a total thickness of 100 to 200 nm (Lipke and Ovalle 1998).

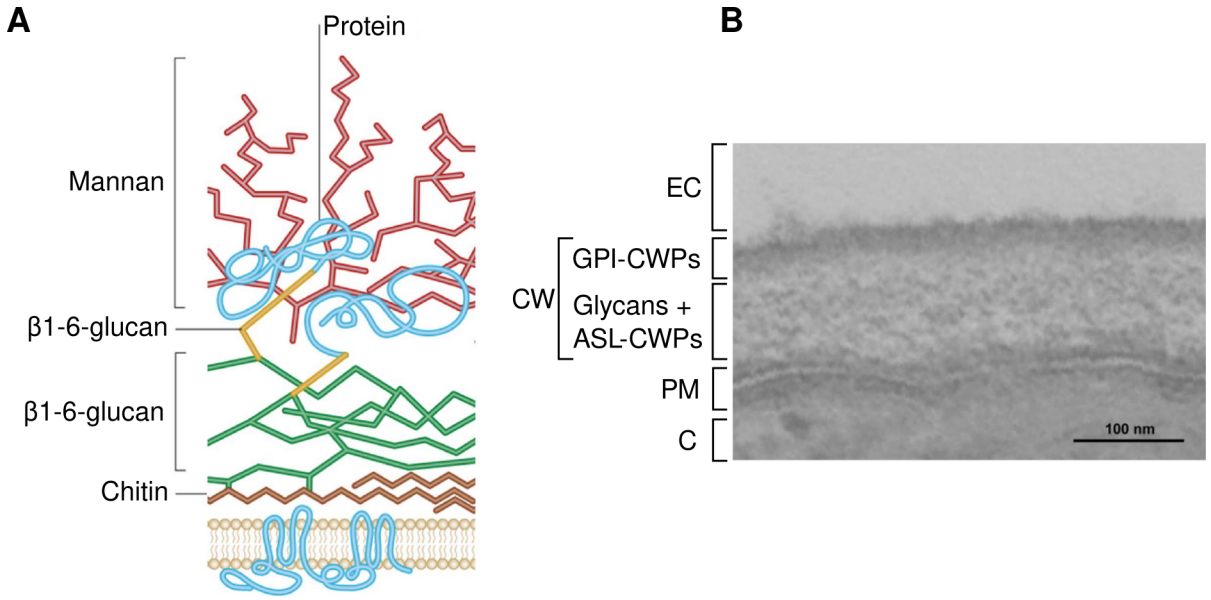

Figure 4: Cell wall structure of diffenrent pathogenic fungi.
A Cartoon representation of the cell wall of *C. glabrata*. A dense network of chitin, branched β1-3 glucan and β1-6 glucan forms the inner cell wall. The outer wall contains highly mannosylated proteins that are anchored to the β1-6 glucan components of the inner cell wall. The figure was adapted from Gow et al. 2017. *B* TEM (transmission electron microscopy) picture of the *C. glabrata* cell wall. The electron-dense outer layer is mainly formed by GPI-CWPs (GPI-modified proteins), that are covalently bound to β1,6-linked glucans. The inner layer of the cell wall is composed of glycans and ASL (alkali-sensitive linkage) proteins that can be released by incubation with low concentrations of NaOH. CW, cell wall; EC, extracellular environment; PM, plasma membrane; C, cytosol. The figure was adapted from de Groot et al. 2008.

The inner layer is composed of β1-3- and β1-6-linked glucans as well as a small amount of chitin (1 to 2 %) that mainly is located at the bud scars (Lesage and Bussey 2006). Interconnection of the three components by extensive cross-linking ensures a sufficient integrity of the cell wall (Kollár et al. 1997). The outer layer of the cell wall consists of mannosylated proteins that form a highly branched network which can make up to 40 % of the cell wall dry weight (Klis et al. 2001). The composition of this outer layer of mannoproteins can vary considerably

depending on species and surrounding conditions as it is constantly modified. Each protein can be highly N- and O-glycosylated, which adds 50 to 100 kDa to the molecular mass of the glycopeptide and produces a thick polysaccharide layer that covers the cell surface (Lipke and Ovalle 1998). A fraction of the glycoproteins is attached to the plasma membrane by a GPI (glycosylphosphatidylinositol) anchor, a posttranslational modification of the C-terminus that can specifically be found in eukaryotic cells (Paulick and Bertozzi 2008). In yeasts, certain proteins further get transferred toward the nonreducing ends of acceptor glycans in the cell wall via transglycosylation of their GPI anchor (Vogt et al. 2020).

To control the synthesis and integrity of this complex structure, a large set of different proteins, enzymes and sensors is necessary. For *S. cerevisiae* it has been shown that of its more than 6000 genes at least 1200 directly or indirectly affect the cell wall which underscores the complexity of the machinery that is needed for maintenance and modification of the cell wall (de Groot et al. 2001). A structurally important group of fungal CWPs are the Pir (proteins with internal repeats) proteins that are covalently connected to β1-3-glucan and thus contribute to the stability of the cell wall (De Groot et al. 2005). They show a uniform distribution throughout the inner cell wall and in case of cell wall damage their production is strongly upregulated (García et al. 2004). To detect cell wall damage and other stresses, sensor proteins are needed that are able to transmit structural information from the cell wall into the cytoplasm to activate a cellular response. One example for a type of sensor proteins is the Wsc protein family. These proteins are suggested to act as mechanosensors and detect mechanical stress or changes in cell wall elasticity. Therefore, these proteins have an extracellular part with an N-terminal cystein-rich domain and a spring-like domain as well as a transmembrane domain and an unstructured C-terminal domain located in the cytoplasm (Kock et al. 2015). The spring-like domain has been postulated to transmit mechanical stress from the cell wall to the cytoplasm where a cellular stress response is triggered (Dupres et al. 2009). By detecting these changes and reacting accordingly, the cell constantly modifies the composition and structure of its cell wall and thus adapts to changes in its environment. While the overall architecture of the cell wall is conserved in many yeasts (Coronado et al. 2007), various species have developed adaptations that are specific for their ecological niche. One adaptation that is generally important for unicellular organisms and pathogens in particular is the attachment to different surfaces. In the case of *C. glabrata* this involves the adhesion to host cells which is essential for its role as a pathogenic organism (de Groot et al. 2013). Adhesion proteins, called adhesins, are relevant virulence factors and thus present possible targets for the development of antifungal drugs which could help to prevent fungal infections in immunocompromised patients.

1.4 Fungal adhesins

The ability to adhere to substrates or other cells is a crucial function of fungal cells that is necessary for the colonization of habitable environments, the formation of biofilms or mating. A number of CWPs (cell wall-attached proteins) has evolved to mediate adhesion to specific targets. In fungi, these adhesins share a common tripartite architecture (Figure 5) that comprises

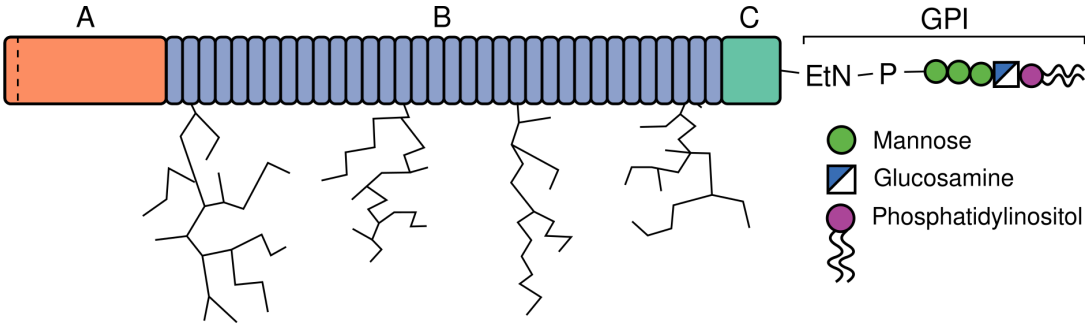

Figure 5: Architecture of fungal adhesins.
Fungal adhesins generally are composed of three protein domains. An N-terminal adhesion domain (A) with a signal peptide (dashed line) is followed by a highly glycosylated central domain (B) consisting of serine and threonine rich repeats. Presumably, glycosylation of the B domain enables a rod-like conformation to effectively present the A domain outside the cell wall (Jentoft 1990). The C-terminal domain harbors a GPI (glycosylphosphatidylinositol) anchor which is composed of ethanolaminephosphate, mannose, glucosamine and phosphatidylinositol.

an N-terminal A (adhesion) domain, a central low complexity B domain and a C-terminal domain that carries a GPI (glycosylphosphatidylinositol) anchor for attachment to the cell wall (Frieman et al. 2002). This anchor first locates the protein to the plasma membrane and subsequently is transferred covalently to the cell wall by specialized transglycosylases (Essen et al. 2020). The adhesion domain is preceded by a signal sequence that directs the protein to the ER, where the B domain, consisting of serine and threonine-rich repeats, is glycosylated. This glycosylation has been proposed to form semi-rigid rod-like superstructures that protrude outside the cell wall and present the adhesion domain (Jentoft 1990). Subsequently, the adhesin is transported to the plasma membrane and eventually transglycosylated to the cell wall. While being crucial for adhesion, the molecular functionality of the N-terminal domain is quite diverse. It can either bind peptides like in the case of the Als (agglutinin-like sequence) protein family of *C. albicans* (Lin et al. 2014), adhere to abiotic surfaces (Brückner and Mösch 2012) or has lectin functionality that allows to specifically bind carbohydrate structures on target surfaces. A lectin functionality was for example found for members of the Flo (flocculin) family of *S. cerevisiae* (Veelders et al. 2010). These proteins confer self recognition and cell-cell adhesion via mannose binding and thus is important for the formation of biofilms and other multicellular aggregates like flocs and flors (Brückner and Mösch 2012). In general, fungal adhesins are best studied in *S. cerevisiae* because of its use in a multitude of industrial applications. Flocculation is for example used in the production of beer, wine, ethanol or biofuel to remove yeast cells after fermentation (Bauer et al. 2010; Soares 2011). Another type of adhesin that has no binding pocket but confers adhesion by homotypic interaction is the flocculin Flo11. Here, two bands of aromatic amino acid residues girdle the adhesion domain and allow specific interaction of identical Flo11A domains (Kraushaar et al. 2015), a feature, that seems to allow kin discrimination within mixed populations of *S. cerevisiae* (Brückner et al. 2020).

A recent analysis revealed that the genome of *C. glabrata* harbors 81 sequences that share the typical architecture of fungal adhesins (Xu et al. 2020). This large number of GPI-CWPs was categorized into seven clusters and twelve singletons by phylogenetic analysis (Figure 6).

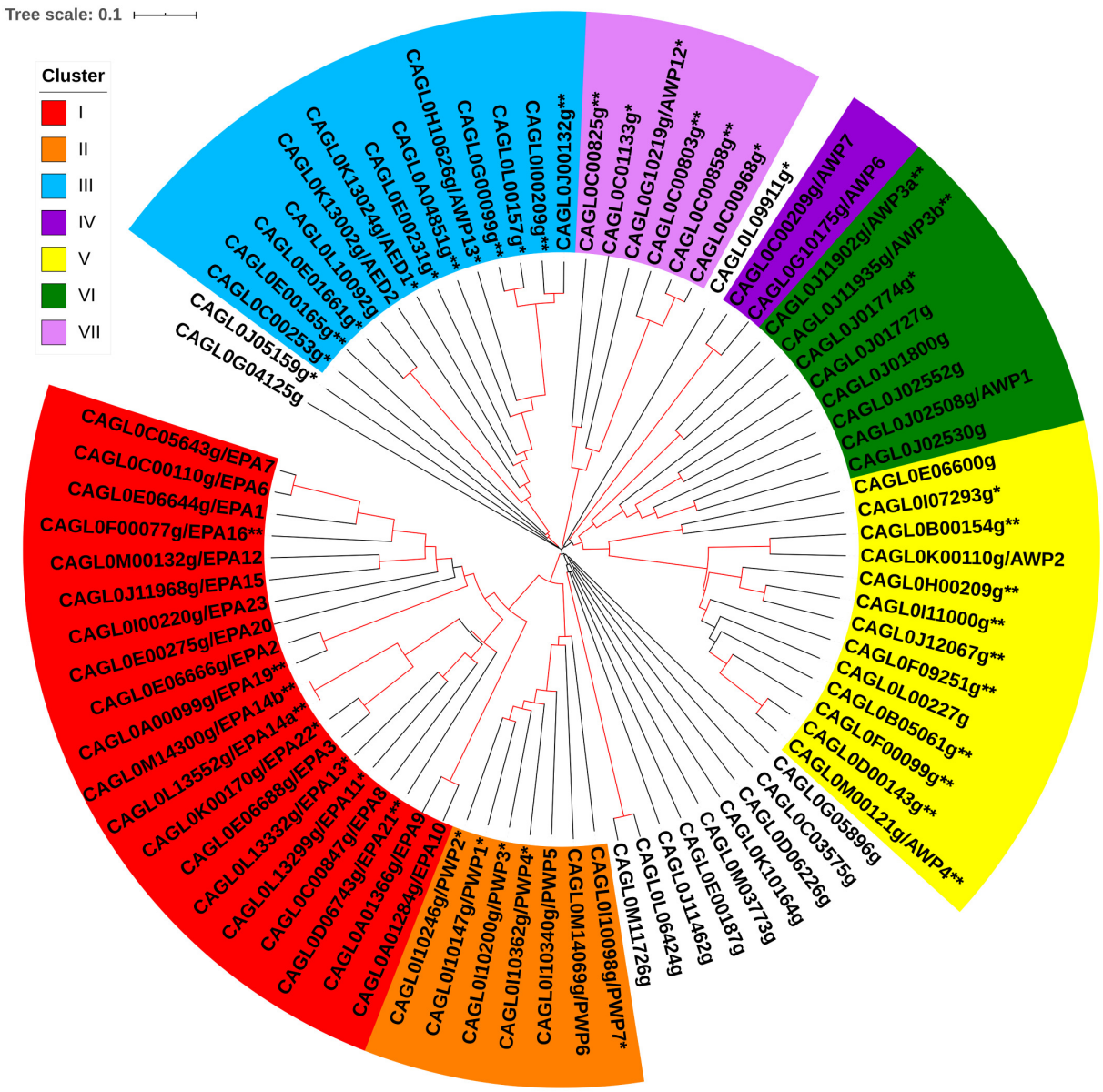

Figure 6: Phylogenetic tree of GPI-anchored adhesins in *C. glabrata*.
The tree was generated by using the N-terminal region of each protein and a bootstrap method with the software ClustalW2 (seed = 111; 1000 bootstrap trials). Genes were clustered according to their phylogenetic relationship and their identifiers were colored according to their respective cluster. Structural variants (*) and novel proteins (**) are indicated. Red branches indicate a bootstrap number >500. The figure was adapted from Xu et al. 2020

The largest cluster is formed by the the Epa (epithelial adhesin) protein family which consists of more than 20 members and represents the best studied group of GPI-CWPs in *C. glabrata*. A second cluster contains the Pwp family, which is closely related to the Epas but only comprises seven members. The other adhesins are largely uncharacterized and have no direct relationship to the Epa or Pwp protein families. Generally, *C. glabrata*'s CWPs are very variable in length with the shortest adhesin having a length of 146 amino acids and the longest consisting of 9859 amino acids (Xu et al. 2020). In total, 19 proteins are longer than 2500 amino acids, but the length does not correlate with cluster membership. About half of all CWPs in *C. glabrata* are located in subtelomeric regions which indicates that their expression usually is epigenetically silenced (De Las Peñas et al. 2003). To date it is not known why *C. glabrata* carries such a

large number of adhesins. As this species can frequently be isolated from different locations, including fermenting coffee beans, bird droppings or mobile phones, it might be possible that a large set of different adhesins enables the distribution of this species over several different habitats (Gabaldón and Fairhead 2019).

1.5 The epithelial adhesin protein family

The Epa protein family is the largest group of GPI-CWPs in *C. glabrata* with 20 members in strain CBS138 (Xu et al. 2020). A varying number of Epa proteins can be found in different members of the *Nakaseomyces* clade with these numbers directly correlating with the pathogenicity of the respective species (Timmermans et al. 2018). A number of studies using either purified adhesion domains or surface display systems has been carried out to investigate the function of Epa proteins and their role as virulence factors. Glycan array analyses of EpaA domains showed that most family members recognize alpha- and beta-linked galactosides, while a smaller fraction binds nongalactosidic terminal glycans (Diderrich et al. 2015). Additionally, *in vivo* adhesion assays with monolayers of human epithelial cells showed that *S. cerevisiae* cells presenting Epa1A, Epa6A or Epa7A can bind very efficiently to host cells. Interestingly, most *EPA* genes found in strain CBS138 are located in subtelomeric regions (Xu et al. 2020) and thus are transcriptionally silenced by chromatin remodeling under laboratory conditions (De Las Peñas et al. 2003; Castaño et al. 2005). As most of the Epa proteins only confer moderate to weak host cell adhesion it seems plausible that these CWPs could provide adhesion to other substrates and enable *C. glabrata* to colonize different habitats. For some family members an additional function besides binding to host cells has already been found. A study on Epa-mediated adhesion to hydrophilic and hydrophobic surfaces could show that *C. glabrata* utilizes Epa1, Epa6 and Epa7 to bind to abiotic substrates (Valotteau et al. 2019). This finding is particularly relevant for clinical environments, as the formation of biofilms on medical devices can promote nosocomial infections, which are a serious problem in modern healthcare (Kojic and Darouiche 2004). Furthermore, Epa8 and Epa19 have been shown to be responsible for adhesion of *C. glabrata* cells to *C. albicans* hyphae in mixed infections (Tati et al. 2016) showing that *C. glabrata* is able to employ various adaptations to switch from its commensal lifestyle to a pathogenic behavior.

Structural analysis of three different EpaA domains revealed several structural features that are shared by all Epa family members. These features are illustrated by a number of high-resolution crystal structures that showed an overall highly conserved PA14/Flo5-like core domain initially found in the protective antigen of the anthrax toxin (Ielasi et al. 2012; Maestre-Reyna et al. 2012; Diderrich et al. 2015; Hoffmann et al. 2020). The PA14 subdomain is part PA20 propeptide of *Bacillus anthracis*, a gram positive bacterium and obligate pathogen. During the maturation of PA20 the PA14 subdomain, named after its molecular weight of 14 kDa, is cleaved off (Petosa et al. 1997). A main characteristic of the PA14 domain is its core domain that consists of a β-sandwich composed of two antiparallel β-sheets (Rigden et al. 2004). Apart from bacterial toxins and yeast adhesins, this fold is also frequently found in glycosidases, glycosyltransferases, proteases and

amidases, showing its wide distribution in different protein classes. This is also reflected by the large number of protein sequences in the Pfam database (https://pfam.xfam.org) belonging to the PA14 superfamily (CL0301). Currently, the database lists 6664 PA14-like sequences from 2236 species (Pfam 33.1, May 2020), including a number of flocculins from *S. cerevisiae*, the Epa family from *C. glabrata* and the chitin binding protein Cea1 of *Komagataella pastoris*.

In the Epa protein family, the conserved core domain is equipped with a number of different structural elements that have been proposed to provide a defined ligand binding affinity and specificity. Two calcium binding loops CBL1 and CBL2 form the inner binding pocket while the outer binding pocket consists of three loops named L1–L3 (Figure 7). Generally, the Epa adhesion domain has a variable surface composition, but certain elements of the inner binding pocket are highly conserved. The group of conserved residues include a D*cis*D motif in CBL1 (Veelders et al. 2010), a tryptophan residue in loop L3 and a disulfide bridge that links L1 and L2 (Figure 7).

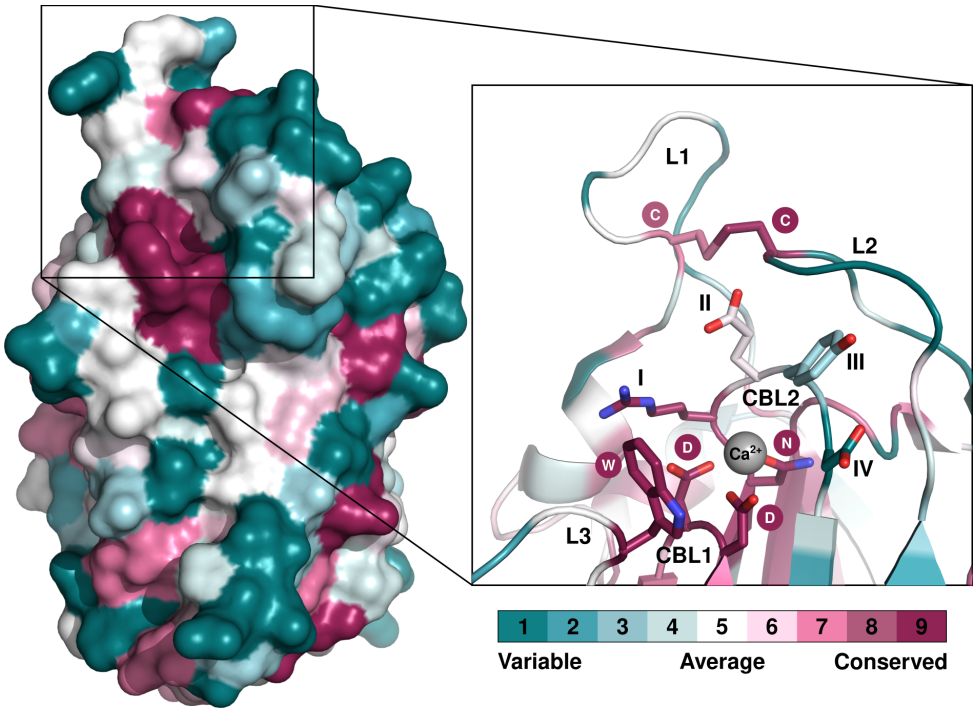

Figure 7: Structural features of EpaA domains
A surface representation of the Epa1A domain is shown on the *left* and was colored according to the conservation of its amino acid residues within the Epa protein family. Color coding was obtained by using the ConSurf server (https://consurf.tau.ac.il/) (Glaser et al. 2003; Landau et al. 2005), the structure of Epa1 (PDB 4A3X) and the multiple sequence alignment shown in Figure 8. The ligand binding pocket is presented on the *right*, showing the conserved D*cis*D motif (CBL1) and an asparagine which are responsible for coordination of an Ca^{2+} ion (grey). Further conserved residues are the tryptophan in loop L3 and two cysteines in loops L1 and L2 forming a disulfide bond. In contrast, positions III and IV of CBL2 are highly variable.

Mutational studies have shown that the conserved structural elements are crucial for ligand binding (Maestre-Reyna et al. 2012; Diderrich 2014). Interestingly, the tryptophan residue in L3 can be substituted with other aromatic residues without any negative impact on functionality. While the inner binding pocket is largely conserved, the outer binding pocket, composed of loops L1 and L2, is highly variable. Additionally, positions II–IV of CBL2 are also highly variable and have been suggested to be involved ligand recognition as their sequence often

correlates with ligand binding specificity. However, as this correlation does not apply for all Epa family members, other variable elements like L1 and L2 were suggested to contribute to ligand recognition (Diderrich et al. 2015). To date, the function of both loops has not been investigated in detail and it remains to be tested to which extent both regions influence ligand binding.

A comparison of all EpaA sequences in strain CBS138 shows that of all structural elements in the adhesion domain the L1 region provides the highest variability (Figure 8). While in most EpaA domains the length of loop L1 is four to eight amino acids, it is significantly elongated in Epa9A and Epa10A which carry L1 regions with a length of 29 or 30 amino acids, respectively. As the putative ligand binding functionality of this region has not yet been addressed by structure-based mutational analysis, it remains unknown whether it is involved in ligand recognition and whether its length directly or indirectly influences ligand binding specificity or affinity. In a previous study, a high-resolution crystal structure of the Epa9A domain was produced to investigate the structural basis for its ligand binding (Diderrich 2014). A comparison of this crystal structure with an existing structure of Epa1A, has shown that despite a moderate sequence identity of 43 % both adhesion domains were structurally nearly identical with an rmsd of 0.64 Å over 144 C_α atoms (Kock 2015). Given this striking structural similarity, an involvement of the variable L1 region in ligand binding might be responsible for the different ligand binding patterns of Epa1A and Epa9A. However, as the crystal structure of Epa9A was lacking a sufficient electron density for the L1 region, a characterization of its structure-function relationship is still missing.

In contrast to the L1 region, the involvement of the CBL2 motif in programming the ligand binding specificity has been investigated in detail. Therefore, a number of studies has been conducted to reveal and modify the ligand binding specificity of the Epa protein family. Glycan array profiling of 17 different adhesion domains of the Epa family was used to categorize them according to their ligand binding patterns (Diderrich et al. 2015). Class I proteins show a preference for glycan structures with β-linked terminal galactosides while class II is composed of family members that preferentially bind α-linked galactosides or sulfated galactosides with β1-3- or β1-4-linkage. The third class comprises the remaining Epa proteins that mainly bind to nongalactosidic glycan structures. Apart from glycan profiling, class II and III both are largely uncharacterized with the exception of Epa6 (class II) that has been analyzed structurally (Diderrich et al. 2015). In addition to functional profiling and structural analysis, the functionality of Epa1A has also been reprogrammed by mutation of the CBL2 motif. It appeared that an exchange of the Epa1A CBL2 motif with corresponding sequences form Epa2A, Epa3A and Epa6A did significantly change its ligand binding pattern and also partly changed its *in vivo* adhesion to human epithelial cells (Maestre-Reyna et al. 2012). Similarly, another approach using random mutagenesis of the CBL2 motif showed a drastic increase in sulfoglycan binding in a number of different mutants (Ielasi et al. 2014). These results show that the CBL2 motif has a central role in defining the ligand binding specificity of EpaA domains but a clear correlation between its composition and the resulting ligand binding patterns has not yet been found.

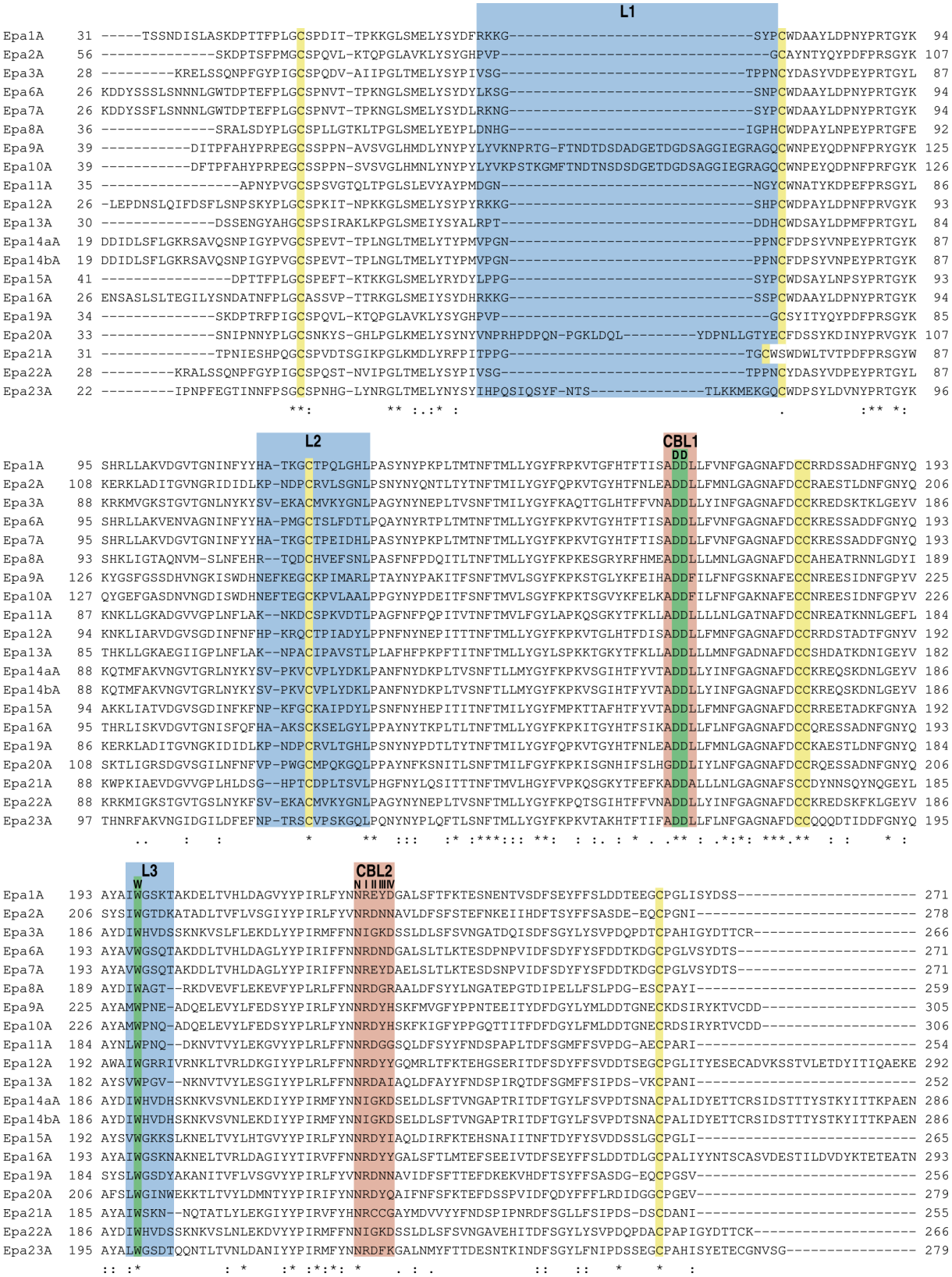

Figure 8: Multiple sequence alignment of EpaA domains
Alignment of all EpaA domains of the *C. glabrata* strain CBS138. The outer binding pocket is composed of the three loops L1, L2 and L3 (blue). CBL1 and CBL2 (red) form the inner binding pocket and are responsible for the coordination of a Ca^{2+} ion and glycan binding. The functionally crucial D*cis*D (DD) motif in CBL1 and the tryptophan (W) residue in L3 are indicated in green. Conservation is indicated below by Clustal Omega consensus symbols.

Of the large Epa family, Epa12A, Epa15A, Epa22A and Epa23A show a preferential binding for sulfated glycans. The same proteins only show a weak to moderate binding to human epithelial cells (Diderrich et al. 2015) leading to the question which target structures are bound by these four adhesins. Sulfated carbohydrates are associated with a large number of different topics like intestinal homeostasis (Bergstrom and Xia 2013), inflammation or cancer growth (Morla 2019). This provides a large number of different targets for adhesins with specific sulfoglycan-binding capabilities. Recently, another family of GPI-CWPs of *C. glabrata*, the PWPs (PA14-containing wall protein), have been shown to potentially bind GAGs (glycosaminoglycans) (Lutterbach 2019). This diverse group of carbohydrates consist of various different disaccharide units which often carry sulfate groups and can be found for example in skin, cartilage and connective tissue (Lee et al. 2016; Zhou et al. 2018). They are also relevant structures for the infections with bacteria, viruses or parasites, which can specifically bind to GAGs during the initial steps of an infection (Rostand and Esko 1997; Bartlett and Park 2010; Green et al. 2013).

1.6 Aims of this study

A first goal of this study was to investigate the role of the variable L1 region in the outer binding pocket of Epa adhesion domains. Therefore, a number of different exchange variants should be created to study the impact of L1 length on *in vivo* and *in vitro* binding behavior. These chimeric exchange variants were then to be functionally characterized by *in vivo* adhesion assays using an *S. cerevisiae* expression system and different human epithelial cell lines. Additionally, complementary *in vitro* measurements should be performed using heterologously produced adhesion domains to determine changes in ligand binding specificity. To accomplish this task, ligand binding patterns should be determined by glycan array experiments and dissociation constants for each variant were to be calculated from fluorescence titration data obtained with mono- and disaccharide ligands. In a next step, a combined exchange of CBL2 motifs and L1 regions should be performed to investigate whether the exchange of both structural elements is sufficient to completely transfer a given ligand binding pattern between donor and recipient domain. A comprehensive understanding on how these two elements determine ligand binding specificity could help to develop lectins with specific binding properties for biotechnological and medical applications. A second aim of this study was to perform a detailed structural characterization of the Epa9A domain in order to elucidate the structure-function relationship between the elongated L1 region and the *in vivo* and *in vitro* binding functionality of the adhesion domain. Therefore, protein crystals of Epa9A were to be soaked with complex branched oligosaccharides to stabilize the conformation of the L1 region and allow the assessment of its structure in complex with a ligand. In a next step, the structures of two CBL2 exchange variants, Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1}, were determined to investigate how the structure of the CBL2 motif controls the ligand binding specificity.

A second part of this study was to investigate the ligand recognition of sulfoglycan-binding EpaA domains, as this topic has not yet been addressed by an in depth analysis. Based on the

available structural information about Epa1A, Epa6A and Epa9A, a number of residues in the CBL2 motifs of Epa12A, Epa15A and Epa23A have previously been identified that putatively are involved in sulfoglycan binding. Based on these data, the three mutants Epa12A^{Y227F}, Epa15A^{Y227F} and Epa23A^{K231A} were to be constructed and characterized by *in vivo* and *in vitro* binding studies. Additionally, a crystallographic analysis of Epa12A and Epa23A should be attempted, as to date no structural information is available for any of the sulfoglycan-binding EpaA domains.

2 Results

2.1 Functional analysis of conserved and variable structural motifs *in vivo*

Epithelial adhesins harbor several conserved and variable structural elements located in the inner and outer binding pocket of their adhesion domains. To address the question which of these elements are directly involved in host cell binding, their functionality has been tested by structure-based mutational analysis. In a previous work, conserved elements of Epa1A have been mutated to create three different variants, Epa1A^{D165A} harboring an A*cis*D motif, Epa1A^{C78S} lacking the disulfide bond between loops L1 and L2 and Epa1A^{W198A} carrying an alanine instead of a tryptophan residue in L3. Additionally, the two mutants Epa1A^{W198H} and Epa1A^{W198Y} were analyzed to test whether the functionality of the tryptophan can be fulfilled by other residues carrying aromatic side chains. These five variants have been tested for *in vivo* adhesion on human epithelial cells using an *S. cerevisiae* expression system and colorectal Caco-2 cells (Diderrich 2014). The adhesion assay revealed that mutation of either the D*cis*D motif, the disulfide bond or W198 results in a significantly reduced host cell adhesion showing that each of the three conserved elements is crucial for functionality. Interestingly, the variants Epa1A^{W198H} and Epa1A^{W198Y} both conferred efficient adhesion indicating that the functionality of W198 can be substituted by other residues with aromatic side chains. In addition to the functionally important conserved elements EpaA domains also harbor variable structural elements such as the CBL2 motif located in the inner binding pocket and loops L1 and L2 that form the outer binding pocket. For Epa1A it has been shown that it is possible to alter ligand binding specificity by either exchange of CBL2 motifs between EpaA variants or random mutagenesis of position II or III in CBL2 (Maestre-Reyna et al. 2012; Ielasi et al. 2014). Subsequently, the functional role of the CBL2 motif was analyzed for other natural EpaA variants, Epa2A, Epa3A, Epa6A and Epa9A, respectively. These experiments could show that in six different chimeric EpaA variants the exchange of a CBL2 motif alters ligand binding specificity but does not transfer it completely (Diderrich 2014).

In this work, the natural Epa10A domain, which has not been characterized before, was tested for host cell adhesion, since Epa10A harbors the same CBL2 motif as Epa9A but exhibits a different ligand binding pattern. Additionally, the two existing CBL2 exchange variants, Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1}, and the newly created mutant Epa10A^{CBL2Epa1} were included to determine the binding behavior of chimeric domains in detail. Parts of the following data were obtained during a master thesis (Friederichs 2018) in the scope of this project. For analysis of *in vivo* binding the adhesion domains were produced in the nonpathogenic *S. cerevisiae* strain BY4741 using a plasmid based expression system (Table 14). Epa adhesion domains were fused to the signal peptide and BC-domain of the flocculin Flo11 of *S. cerevisiae* to ensure correct localization and presentation on the cell surface (Figure 9A). Expression and localization of all constructs was tested by immunofluorescence microscopy using an anti-HA antibody. The

average fluorescence intensity was quantified relative to Epa1A using at least three clones of each strain derived from three independent transformation events. Here, all strains produced sufficient amounts of fusion protein and presented it on the cell surface (Figure 9B).

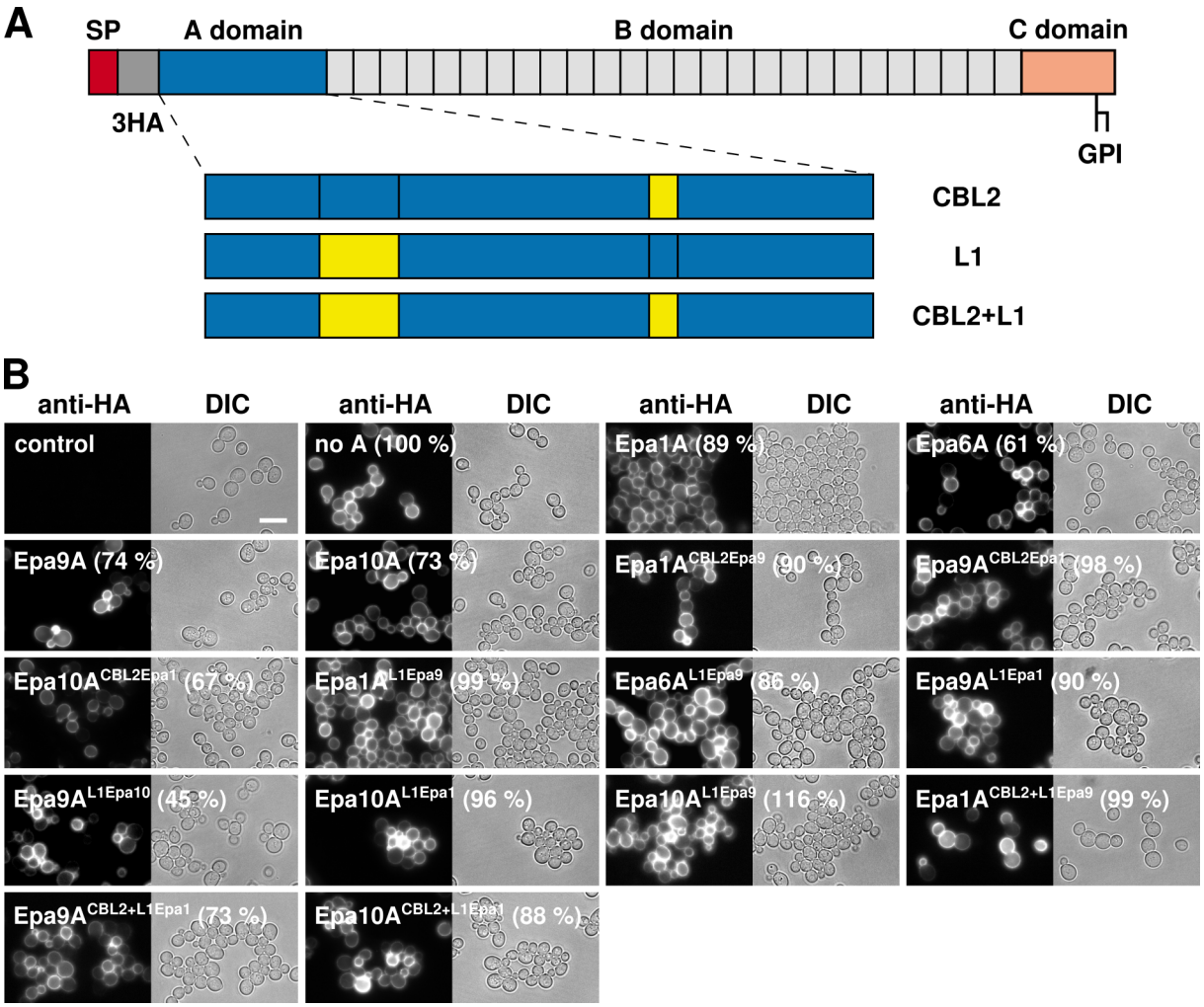

Figure 9: Domain architecture and *in vivo* localization of EpaA variants characterized in this work.
A. The *EPAA* expression cassettes used in this work contained the signal peptide (SP, red), B domain (light grey) and C domain (light red) of the *FLO11* flocculin from *S. cerevisiae*. For immunofluorescence microscopy a 3HA tag (grey) was inserted between SP and *EPAA* domain (blue). Chimeric adhesion domains were constructed by exchange of either the CBL2 motif, the L1 region or both (yellow). *B*. Expression and localization of fusion constructs were tested by fluorescence microscopy using anti-HA primary and Cy3-conjugated secondary antibodies. Labeled cells were visualized and detected using differential interference contrast (DIC) and a rhodamine filter set (anti-HA). An empty vector and an expression construct without an adhesion domain (no A) were used as controls. Fluorescence intensities are shown relative to Epa1A. Scale bar corresponds to 10 µm.

In vivo host cell binding was then determined by adhesion assays on monolayers of human epithelial cells. For this purpose, the Caco-2 cell line was used which is derived from a human colorectal adenocarcinoma and was identified to be suited for adhesion experiments with *Candida* species (Dieterich et al. 2002). Here, the natural variants Epa1A and Epa6A showed a relative target cell adhesion of 59 % and 49 %, while Epa9A and Epa10A exhibited weaker host cell binding of 22 % and 21 % (Figure 10). These results correspond to the adhesion of Epa1A, Epa6A and Epa9A found in other studies (Maestre-Reyna et al. 2012; Diderrich et al. 2015).

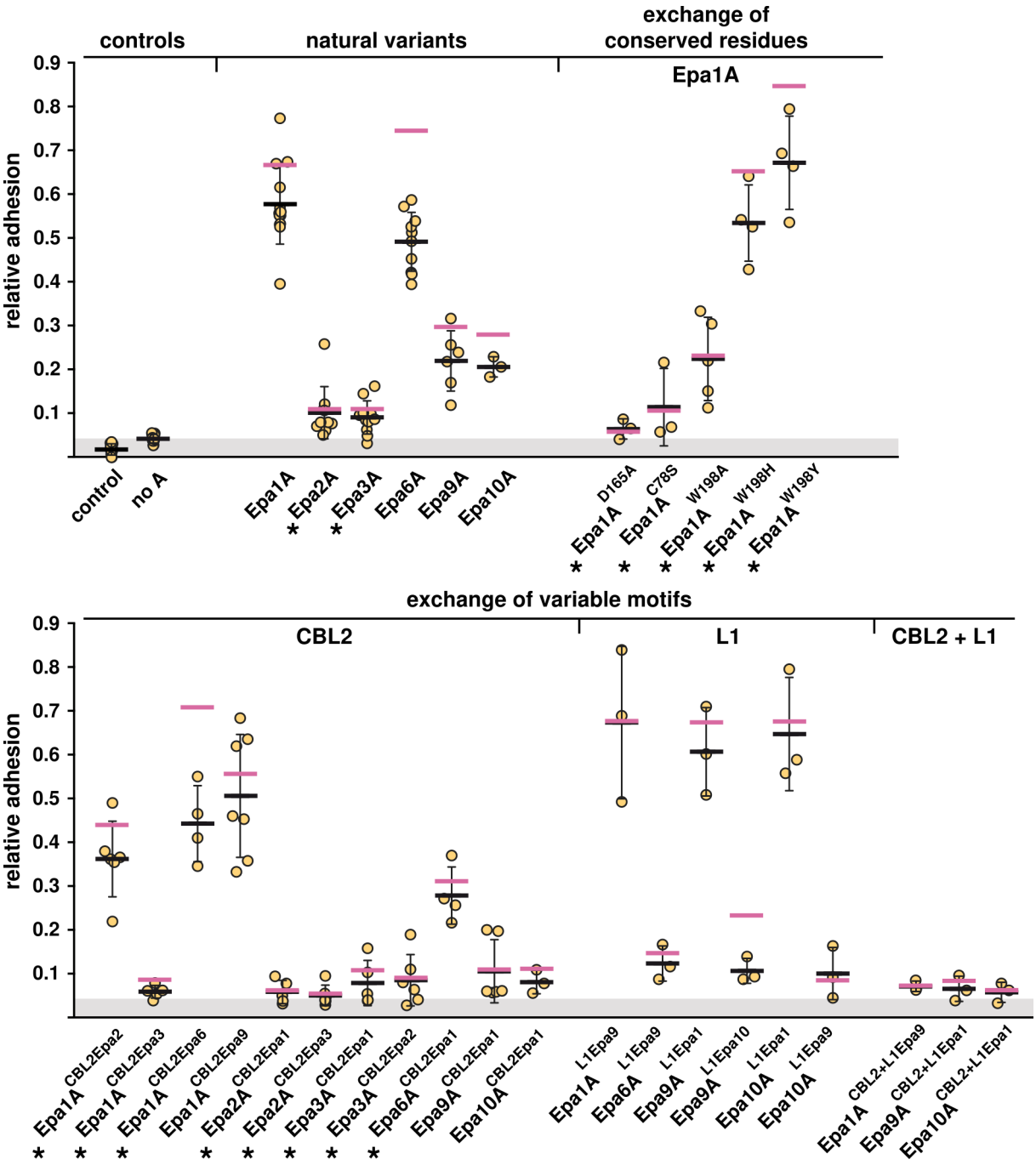

Figure 10: *In vivo* binding of EpaA domains to human epithelial cells.
Host cell adhesion of different exchange variants was tested by using the *S. cerevisiae* strain BY4741 carrying appropriate plasmids shown in Figure 9B and monolayers of the human epithelial cell line Caco-2. Plasmids carried either no adhesin (control), an expression cassette without an adhesion domain (no A) or different natural or mutant EpaA variants as indicated. Yeast strains were grown to logarithmic phase and labelled by using Cy3 conjugated antibodies. Host cell adhesion was then determined after incubation for 2 h on Caco-2 monolayers. Therefore, nonadherent yeast cells were removed by washing and fluorescence of adhesive cells was measured by fluorimetry. Relative adhesion was determined by calculating the ratio between adhesive cells and the total amount of cells used in the assay. At least three independent measurements (yellow dots) were used for determination of average adhesion (black bars) and standard deviation (error bars). Magenta bars represent the average adhesion corrected for the amount of EpaA domains on the cell surface given in Figure 9B. Nonspecific adhesion resulting from the carrier domain is shown by a grey band. Statistical significance was calculated by an unpaired *t*-test resulting in *p* values given in Table 1. These values were used for classification of EpaA domains relative to their respective recipient and donor domains. Adhesion data were partly obtained by Sabrina Friederichs during a master thesis in the scope of this project. Data for additional EpaA variants (Diderrich 2014) are marked by an asterisk and were included to allow direct comparison with results obtained in this work.

In this work, all chimeric EpaA variants were further categorized according to their *in vivo* binding behavior into three groups that could resemble either the functionality (i) of the recipient, (ii) of the donor, or (iii) novel characteristics not observed in the respective recipient or donor variant. For direct comparison with previously described CBL2 exchange mutants, their host cell adhesion data were added to Figure 10. This comparison reveals that the four chimeric variants Epa1A^{CBL2Epa9}, Epa2A^{CBL2Epa1}, Epa3A^{CBL2Epa1} and Epa9A^{CBL2Epa1} exhibit an *in vivo* binding behavior that is similar to their respective recipient variants. Thus, the high efficiency binding of Epa1A or the low efficiency binding of Epa2, Epa3 or Epa9 was not affected by exchange of the CBL2 motif. A change of binding behavior toward a donor variant was only found in one case, Epa1A^{CBL2Epa3}, which showed a significantly lower adhesion strength, similar to Epa3A. The three chimeras Epa1A^{CBL2Epa2}, Epa6A^{CBL2Epa1} and Epa10A^{CBL2Epa1} exhibited a binding pattern that was comparable neither to the recipient nor to the donor and thus was classified as novel. For the three variants Epa1A^{CBL2Epa6}, Epa2A^{CBL2Epa3} and Epa3A^{CBL2Epa2} the observed binding efficiency was similar to both, recipient and donor. In general, these results show that EpaA mediated host cell adhesion could principally be reprogrammed by the exchange of CBL2 motifs. However, it was not possible to selectively transfer a specific binding pattern between donor and recipient, because the CBL2 exchange variants exhibited recipient, donor and novel behavior to a comparable extent. This analysis suggests that other variable regions must be involved in programming binding behavior of EpaA domains.

To investigate the involvement of other structural elements in host cell adhesion and ligand binding, the L1 region was chosen, because it is highly variable in sequence composition and length Figure 8. Furthermore, previous studies have suggested that the L1 region could be responsible for fine-tuning of the binding pocket by indirectly influencing the conformation of the inner binding pocket (Diderrich et al. 2015). To address this issue, six different L1 exchange variants were constructed based on the natural variants Epa1A, Epa6A, Epa9A and Epa10A. These natural variants were chosen because Epa1A, Epa6A and Epa9A are structurally well characterized which allows a solid design of mutants. Furthermore, Epa9A and Epa10A carry the same CBL2 motif but differ in ligand binding specificity. Additionally, both EpaA variants harbor an L1 region which is considerably longer than that of other EpaA domains with a total length of 29 or 30 amino acids, respectively. Of the six tested L1 exchange mutants, Epa1A^{L1Epa9} conferred target cell adhesion with a binding strength comparable to its recipient variant. In the case of Epa9A^{L1Epa10} and Epa10A^{L1Epa9}, no significant difference could be found toward their respective donor and recipient variant. For Epa6A^{L1Epa9}, host cell binding was reduced to 13 % and the two variants Epa9A^{L1Epa1} and Epa10A^{L1Epa1} showed a significant increase (Table 1) in adhesion and now exhibit target cell binding comparable to Epa1A. Thus, these three chimeras exhibit the behavior of their respective donor variants.

To address the question, whether a combined transfer of CBL2 and L1 is sufficient to completely transfer *in vivo* binding behavior of EpaA domains, the two double exchange mutants Epa9A$^{CBL2+L1Epa1}$ and Epa10A$^{CBL2+L1Epa1}$ were tested. Both mutants showed low host cell

Table 1: Statistical significance for comparative analysis of effects of mutations in EpaA variants on epithelial cell adhesion (data shown in Figure 10)

	***p*-values[1] for comparison to**		
EpaA variant	**Recipient**	**Donor**	**Other**
Epa1A	< 0.0001	-	-
Epa2A	0.003	-	n.s.[2] vs[3] Epa3A
Epa3A	0.0003	-	n.s. vs Epa2A
Epa6A	< 0.0001	-	-
Epa9A	< 0.0001	-	n.s. vs Epa10A
Epa10A	< 0.0001	-	n.s. vs Epa9A
Epa1A^{D165A}	< 0.0001	-	-
Epa1A^{C78S}	< 0.0001	-	-
Epa1A^{W198A}	< 0.0001	-	-
Epa1A^{W198H}	n.s.	-	-
Epa1A^{W198Y}	n.s.	-	-
Epa1A^{CBL2Epa2}	< 0.0001	< 0.0001	-
Epa1A^{CBL2Epa3}	< 0.0001	n.s.	-
Epa1A^{CBL2Epa6}	n.s.	n.s.	-
Epa1A^{CBL2Epa9}	n.s.	0.0008	-
Epa2A^{CBL2Epa1}	n.s.	< 0.0001	-
Epa2A^{CBL2Epa3}	n.s.	n.s.	-
Epa3A^{CBL2Epa1}	n.s.	< 0.0001	-
Epa3A^{CBL2Epa2}	n.s.	n.s.	-
Epa6A^{CBL2Epa1}	0.0002	< 0.0001	-
Epa9A^{CBL2Epa1}	n.s.	< 0.0001	-
Epa10A^{CBL2Epa1}	0.004	< 0.0001	-
Epa1A^{L1Epa9}	n.s.	0.0006	-
Epa6A^{L1Epa9}	< 0.0001	n.s.	-
Epa9A^{L1Epa1}	0.0002	n.s.	-
Epa9A^{L1Epa10}	n.s.	n.s.	-
Epa10A^{L1Epa1}	0.004	n.s.	-
Epa10A^{L1Epa9}	n.s.	n.s.	-
Epa1A$^{CBL2+L1Epa9}$	< 0.0001	0.009	0.0008 vs Epa1A^{CBL2Epa9} 0.004 vs Epa1A^{L1Epa9}
Epa9A$^{CBL2+L1Epa1}$	0.008	< 0.0001	n.s. vs Epa9A^{CBL2Epa1} 0.0009 vs Epa9A^{L1Epa1}
Epa10A$^{CBL2+L1Epa1}$	0.0013	< 0.0001	n.s. vs Epa10A^{CBL2Epa1}

[1] The *p*-values were calculated by applying an unpaired *t*-test using the data present in Figure 10.

[2] Differences between compared variants are rated as not significant (n.s.) for P values exceeding 0.01.

[3] vs = versus.

adhesion comparable with their corresponding CBL2 exchange variants. This suggests that the exchange of the CBL2 motif has a more pronounced effect on target cell binding than an exchange of the L1 region. In case of the third double exchange mutant, Epa1A$^{CBL2+L1Epa9}$, a novel adhesion behavior was found. This result is especially striking, because the Epa1A^{CBL2Epa9} and Epa1A^{L1Epa9} single exchange variants both exhibit behavior of the Epa1A recipient, which shows that in the case of Epa1A it is not possible to reprogram its binding behavior by exchanging only one structural motif.

In summary, the results of the *in vivo* adhesion assay suggest that both, the CBL2 motif and the L1 region, play an important role in host cell adhesion. The exchange of the variable CBL2 motif further reprograms the ligand binding specificity of EpaA domains without transferring the donor binding pattern. Instead, the chimeric EpaA variants exhibit novel specificities and only partly resemble the binding behavior of their respective recipients. In contrast, the exchange of loop L1 principally can transfer the ability to strongly bind epithelial cells between EpaA domains but generally does not change their ligand binding specificities. However, in the case of Epa1A this behavior cannot be reproduced as the exchange of the L1 region did not significantly alter host cell binding of this variant. Furthermore, a combined exchange of L1 and CBL2 abrogated the adhesion gain of the L1 exchange and resulted in weak host cell binding.

In a next step, the *in vivo* adhesion assays were repeated with two additional human cell lines to investigate whether binding patterns found with Caco-2 cells can be validated with cells that represent other sites of the human body. Therefore, the TCC-SUP and HeLa cell lines were chosen which are derived from tumor cells that were extracted from the human bladder and genital tract, respectively, and thus represent other common infection sites for *C. glabrata*. A direct comparison of cellular growth showed minor differences in the formation of monolayers in the three compared cell lines. Caco-2 and TCC-SUP both formed smooth monolayers while in the case of HeLa cells small multilayered patches could be found. In the other two cell lines this only occurred after reaching full confluence. Generally, EpaA mediated adhesion to TCC-SUP cells was lower than to Caco-2 cells but showed a comparable pattern with an overall strong adhesion for Epa1A, Epa6A, Epa1A^{CBL2Epa9}, Epa1A^{L1Epa9}, Epa9A^{L1Epa1} and Epa10A^{L1Epa1} (Table 2). In case of the HeLa cell line adhesion strength was very similar to Caco-2 for most of the tested EpaA variants but in general these results showed a higher standard deviation. This most likely results from nonspecific binding of yeast cells indicated by a relative adhesion of 11 to 13 % for the controls. In summary the adhesion pattern for the tested EpaA variants was comparable on all three human epithelial cell lines. Additionally, the exchange of the CBL2 motif and L1 region showed no cell line specific effect on host cell binding.

Table 2: Relative adhesion of different EpaA variants on human epithelial cell lines.

Variant	Caco-2 (%)	TCC-SUP (%)	HeLa (%)	Behavior
control	2 ± 1	1 ± 1	11 ± 10	natural
no A	4 ± 1	1 ± 1	13 ± 8	natural
Epa1A	59 ± 9	44 ± 7	54 ± 3	natural
Epa1A^{CBL2Epa9}	51 ± 13	43 ± 7	52 ± 8	recipient
Epa1A^{L1Epa9}	67 ± 14	42 ± 4	56 ± 3	recipient
Epa1A$^{CBL2+L1Epa9}$	7 ± 1	2 ± 1	20 ± 15	novel
Epa6A	49 ± 6	29 ± 3	41 ± 6	natural
Epa6A^{L1Epa9}	13 ± 3	2 ± 2	19 ± 13	donor
Epa9A	22 ± 6	4 ± 3	9 ± 6	natural
Epa9A^{CBL2Epa1}	11 ± 7	1 ± 1	9 ± 8	novel
Epa9A^{L1Epa1}	61 ± 8	41 ± 6	56 ± 7	donor
Epa9A$^{CBL2+L1Epa1}$	7 ± 2	2 ± 2	19 ± 15	novel
Epa9A^{L1Epa10}	11 ± 2	2 ± 2	18 ± 13	recipient/donor
Epa10A	21 ± 2	3 ± 2	15 ± 10	natural
Epa10A^{CBL2Epa1}	8 ± 2	2 ± 1	14 ± 8	novel
Epa10A^{L1Epa1}	65 ± 11	31 ± 6	50 ± 5	donor
Epa10A$^{CBL2+L1Epa1}$	6 ± 2	2 ± 1	11 ± 9	novel
Epa10A^{L1Epa9}	12 ± 10	3 ± 2	15 ± 10	recipient/donor

2.2 Functional analysis of variable structural motifs *in vitro*

2.2.1 Production and purification of CBL2 and L1 exchange variants

After the characterization of *in vivo* host cell binding on human epithelial cells, the exchange variants given in Table 2 were further analyzed *in vitro* by determination of ligand binding patterns. Therefore, adhesion domains of chimeric EpaA variants first were heterologically produced using the pET-28(a)+ expression system and the *E. coli* strain SHuffle T7 Express. This strategy has been successfully used for production of a number of fungal cell wall proteins including the *S. cerevisiae* adhesion domains Flo5A, Flo11A and a number of natural EpaA variants (Veelders et al. 2010; Maestre-Reyna et al. 2012; Kraushaar 2016). With this method, soluble protein could be obtained for the three CBL2 mutants Epa1A^{CBL2Epa9}, Epa9A^{CBL2Epa1} and Epa10A^{CBL2Epa1}, the five L1 exchange variants Epa1A^{L1Epa9}, Epa6A^{L1Epa9}, Epa9A^{L1Epa1}, Epa9A^{L1Epa10} and Epa10A^{L1Epa9} and the three double exchange mutants Epa1A$^{CBL2+L1Epa9}$, Epa9A$^{CBL2+L1Epa1}$ and Epa10A$^{CBL2+L1Epa1}$. Purification of natural and chimeric EpaA domains was performed by a combination of IMAC (immobilized metal ion affinity chromatography) and SEC (size exclusion chromatography). To prevent interaction of EpaA domains with the column material during purification, lactose was added to the buffers used for IMAC and SEC. Purified

proteins were then used for further characterization by glycan array analysis, fluorescence titration spectroscopy and crystallization (Table 3). The three mutants Epa1A^{CBL2Epa6}, Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} have initially been described by Maestre-Reyna et al. (2012) and Rike Diderrich (2014). The construction of double exchange variants Epa1A$^{CBL2+L1Epa9}$, Epa9A$^{CBL2+L1Epa1}$ and Epa10A$^{CBL2+L1Epa1}$ was carried out by Sabrina Friederichs (2018) during a master thesis in the scope of this work.

Table 3: Protein yield and experiments performed with CBL2 and L1 exchange variants

Variant	Protein yield (mg/l)	Glycan array	Crystallization	K_D
Epa1A^{CBL2Epa6}	2	✓	-	-
Epa1A^{CBL2Epa9}	1	✓	✓	✓
Epa9A^{CBL2Epa1}	2	✓	✓	✓
Epa1A^{L1Epa9}	5	✓	-	✓
Epa6A^{L1Epa9}	1	✓	-	✓
Epa9A^{L1Epa1}	5	✓	-	✓
Epa9A^{L1Epa10}	5	✓	-	✓
Epa10A^{L1Epa9}	1	✓	-	✓
Epa1A$^{CBL2+L1Epa9}$	1	-	-	✓
Epa9A$^{CBL2+L1Epa1}$	1	-	-	✓

2.2.2 Glycan array analysis of L1 exchange variants

In this work, it was shown that the L1 region of the epithelial adhesion domains is involved in *in vivo* binding to human epithelial cells (Figure 10). However, it is not known whether loop L1 affects the ligand binding specificity or affinity or both. To address this question, five purified EpaA domains carrying exchanged L1 regions were fluorescently labeled with Alexa Fluor 488 TFP ester and sent to the Consortium for Functional Glycomics (CFG) for glycan array analysis. The labeled adhesion domains were tested on mammalian glycan array chips (v5.4) harboring 585 different glycan structures which resulted in specific binding patterns for each tested adhesion domain (Figure 11). Raw fluorescence values were normalized to the best bound glycan for each EpaA variant and analyzed by cluster analysis and Terminal Disaccharide Analysis (TDA). Therefore, available array data of natural and CBL2 exchange variants (Diderrich 2014) were included which resulted in a set of 18 different adhesion domains. To date no cluster analysis of CBL2 or L1 exchange variants has been performed. In a first step, all glycan array data used in this analysis were adjusted to the number of glycans present on mammalian glycan array chips v5.4. Then, hierarchic clustering was performed with the Pearson correlation method which evaluates the linear relationship between two sets of data.

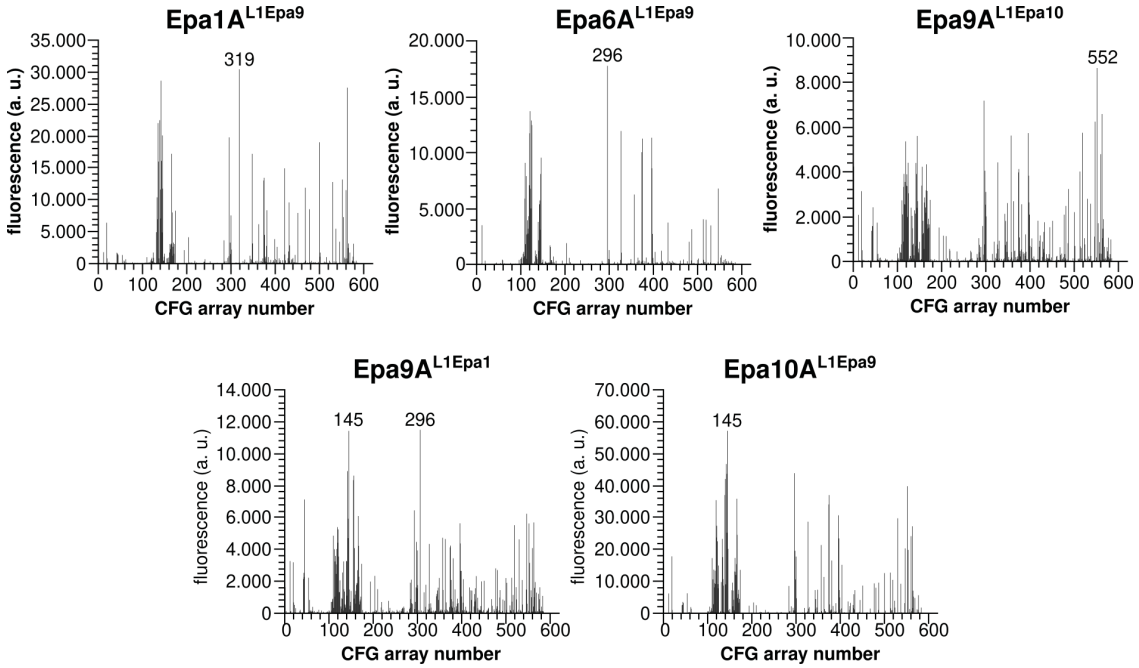

Figure 11: Glycan binding profiles of L1 exchange variants constructed in this study.
Binding profiles for L1 exchange variants were obtained by using mammalian glycan arrays (version 5.4) of the Consortium for Functional Glycomics (CFG). EpaA proteins were purified from *E. coli* and fluorescently labeled with Alexa Fluor 488 TFP Ester before glycan profiling. Fluorescence units measured for different Epa variants are shown by black lines. The best bound glycans are indicated by their CFG glycan ID number.

A linear relationship is defined as a change in one variable with a proportional change in the other variable. The output is visualized in Figure 12 by a color matrix showing the relative binding strength with values from 0 (yellow) to 100 (blue). Additionally, all EpaA variants and glycan structures were sorted by pair-wise distance of binding patterns which is depicted by dendrograms to the left and on top of the matrix, respectively. The cluster analysis revealed very similar ligand binding patterns for the four variants Epa1A^{L1Epa9}, Epa6A^{L1Epa9}, Epa9A^{L1Epa1} and Epa9A^{L1Epa10} and their respective recipient variants Epa1A, Epa6A and Epa9A. Only one exception, Epa10A^{L1Epa9}, could be found to cluster closer to the donor than to the recipient. According to the matrix in Figure 12, this variant seems to retain the specificity of the recipient but also gain additional specificity that is similar to the donor, Epa9A. In case of the seven CBL2 exchange variants it was found that one, Epa9A^{CBL2Epa1}, showed a ligand binding pattern that was similar to the donor. Here, the mutant appears to have a narrowed specificity which is shared by both, the donor and the recipient and in general exhibits a very weak ligand binding. For the other six variants, a novel binding pattern was found. Of these six, Epa1A^{CBL2Epa2}, Epa1A^{CBL2Epa3} and Epa1A^{CBL2Epa6} showed a similar ligand binding pattern that largely lacks specificity shown by the recipient. Instead, these three variants gained specificity for ligands that are also bound by Epa12A and Epa22A. Two other variants, Epa1A^{CBL2Epa9} and Epa3A^{CBL2Epa1}, also gained novel specificity, but appear to partially retain the ligand binding behavior of Epa1A and Epa3A. The five variants Epa1A^{CBL2Epa2}, Epa1A^{CBL2Epa3}, Epa1A^{CBL2Epa6}, Epa1A^{CBL2Epa9} and Epa3A^{CBL2Epa1} all clustered close to Epa12A, Epa15A, Epa22A and Epa23A which have been described to specifically bind sulfated galactosides (Diderrich et al. 2015).

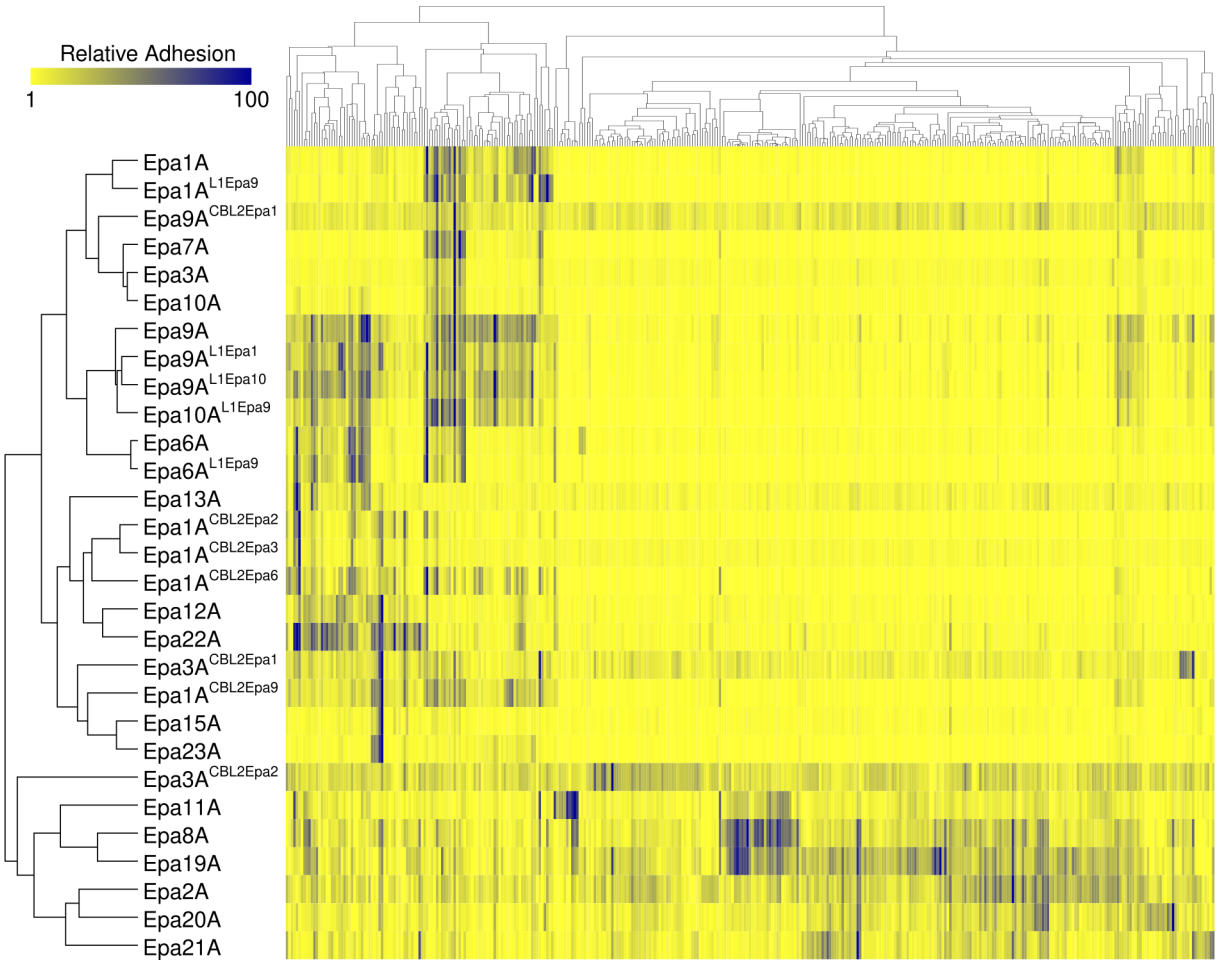

Figure 12: Cluster analysis of glycan array data for different exchange variants.
Binding patterns of chimeric Epa adhesion domains were analyzed by cluster analysis with the Morpheus tool (https://software.broadinstitute.org/morpheus) using a hierarchical approach (Pearson method). Relative *in vitro* binding strength of purified EpaA L1 exchange domains was determined by glycan array analysis with 585 different glycans present on the CFG array version 5.4. Raw data of all available natural and CBL2 exchange variants were first normalized to the best bound ligand for each EpaA variant and then adjusted to array version 5.4. Dendrograms on the left and on top of the matrix were created by pairwise comparison of rows and columns, respectively, and represent similarity of binding patterns. Relative binding strength is given by a color scale from 0 (yellow) to 100 (blue). A list of all glycan structures present on chip version 5.4 is shown in Table A2.

In the case of Epa3A$^{\text{CBL2Epa2}}$ the ligand binding specificity was shifted completely and did not resemble any of the previously describe EpaA variants while also showing nonspecific binding of the glycan structures present on the chip. The cluster analysis therefore placed this variant near a group of EpaA variants at the bottom of the matrix that generally showed a high amount of nonspecific binding.

In summary, the exchange of an L1 region between two EpaA domains does not substantially change the ligand binding behavior of the recipient. In contrast, the CBL2 exchange variants largely show an altered ligand binding pattern which in five of six cases leads to a novel specificity. Furthermore, in two cases an increased amount of nonspecific binding could be found. This suggests, that the L1 region is not critical for determination of ligand binding specificity whereas the CBL2 seems to be directly involved in programming specificity.

For a more detailed analysis of ligand binding specificity, all EpaA variants were categorized according to their binding to terminal disaccharide types. Therefore, the terminal and penultimate carbohydrate moiety of all bound glycans were taken into account. Pie charts shown in Figure 13A represent all glycan structures with unambiguous terminal disaccharide types and a relative binding of at least 50 %. The chart areas correspond to the logarithm of the fluorescence signal over noise for the best bound ligand. Natural and CBL2 exchange variants have been described before (Diderrich 2014; Diderrich et al. 2015) and were adjusted to the number of glycan structures present on chip v5.4.

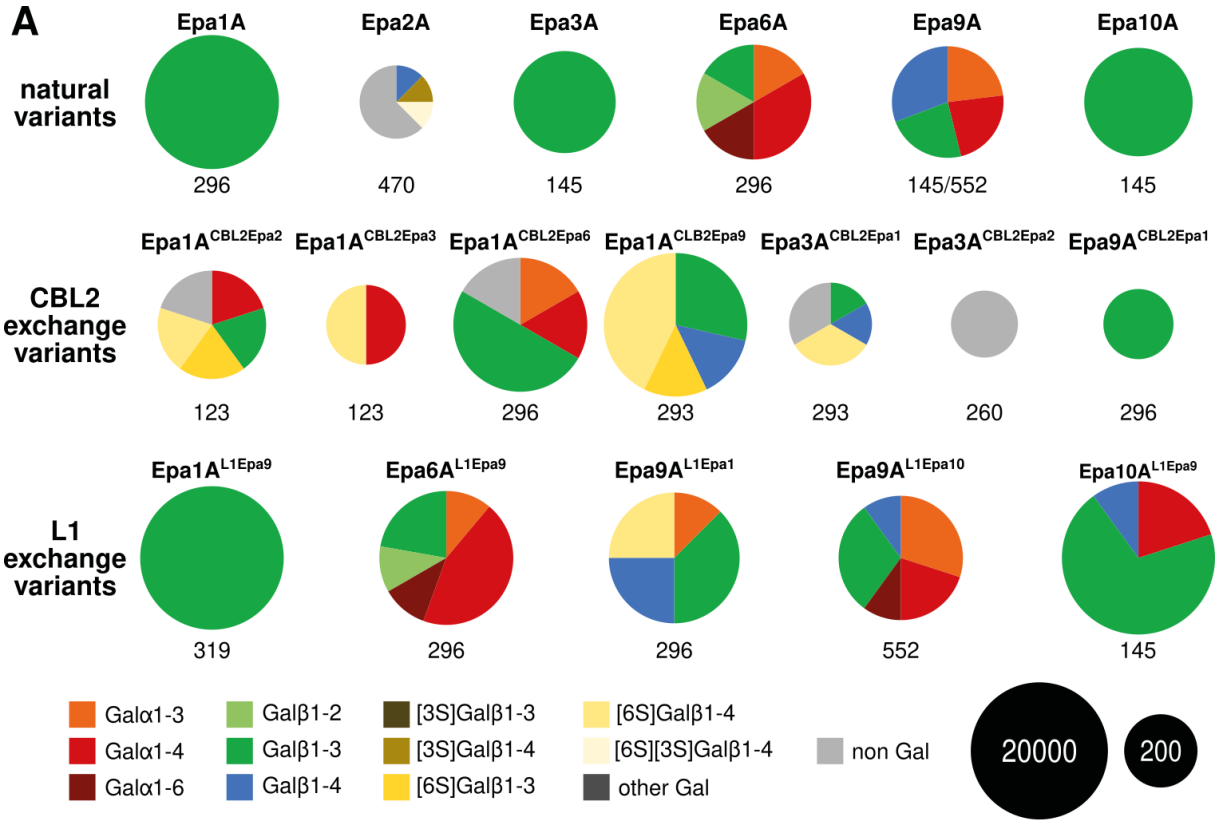

Figure 13: Terminal disaccaride analysis of chimeric EpaA variants.
A. Ligand binding pattern of different EpaA variants is represented by pie charts. Color coding indicates the respective categories of glycans with unambiguous terminal sugar moieties and a relative binding of at least 50 %. Ligand binding pattern for natural variants and CBL2 exchange variants have been described before (Diderrich 2014; Diderrich et al. 2015) and were adjusted to match the number of glycans on the mammalian glycan array chip v5.4 used for the L1 exchange variants. B. Structures of the best bound glycan structures are indicated by glycan ID for each variant and represented according to the Symbol Nomenclature for Glycans (Varki et al. 2015). Chart sizes correnspond to the logarithm of the fluorescence signal over noise for the best bound ligand on the respective chip.

This allows a direct comparison of novel specificity profiles with the existent data. The best bound ligand for each EpaA variant is indicated and the corresponding glycan structure is shown in Figure 13B according to the Symbol Nomenclature for Glycans (Varki et al. 2015). For Epa1A^{L1Epa9} and Epa6A^{L1Epa9} the overall ligand binding pattern was comparable to that of Epa1A and Epa6A respectively, but a new best bound ligand (319) was found for Epa1A^{L1Epa9}. Ligand 319 is a bi-antennary galactoside of the same glycan category as the best bound ligand of Epa1A. Both belong to the category of β1-3 linked galactosides. In the cases of Epa6A^{L1Epa9} and Epa9A^{L1Epa10}, the ligand binding profiles were nearly identical with their recipient domains, Epa6A and Epa9A. Additionally the chimeric variants also had the same best bound ligands as their recipients. For the L1 exchange variant Epa9A^{L1Epa1}, a pattern was found that was similar to the recipient, but contained additional specificity for sulfated, β1-4 linked galactosides. Another variant with additional specificity but otherwise unchanged ligand binding profile is Epa10A^{L1Epa9}. Here, in addition to the recipient's specificity for β1-3 linked galactosides, also α1-3 and β1-4 linked galactosides are bound. In general, all five L1 exchange variants showed ligand binding patterns that were similar to the respective recipients, which is in contrast to results obtained for the CBL2 exchange variants. An exchange of CBL2 motifs always resulted in binding patterns with clearly different specificity.

In addition to the mammalian printed array of the CFG, a second glycan array chip was used for further analysis of chimeric EpaA domains. This analysis was performed in cooperation with the group of Prof. Dr. Carlo Unverzagt at the University of Bayreuth (Germany). The new glycan array chip contained 80 novel carbohydrate structures that have not been present on previously used chips. Most glycans present on the new chip were complex carbohydrates with two or more terminal sugar moieties. A set composed of three natural (Epa1A, Epa6A and Epa9A) and three chimeric EpaA variants (Epa1A^{CBL2Epa6}, Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1}) was chosen for analysis with the new chip (Figure 14A). The three CBL2 exchange variants were chosen because they represent two different cases in which the ligand binding specificity changed toward the donor (Epa9A^{CBL2Epa1}) or was different to both donor and recipient behavior. For the three different Epa1A variants a preference for glycans with at least one terminal β1-4 linked galactose was determined (Figure 14B). Fluorescence signals of Epa9A, Epa6A and Epa9A^{CBL2Epa1} were very weak but also indicated binding to Galβ1-4 which here is about 100 times lower than for Epa1A. Interestingly, the only β1-3-linked galactoside on the chip, Galβ1,3-GlcNAcβ1,4-Galβ1,4-GlcNAc, is not bound by any of the tested EpaA variants. This result is surprising because this particular glycan is very similar to glycan 145 (Galβ1-3GalNAcβ1-4Galb1-4Glcβ-Sp8) of the CFG glycan array which there was the best binder for Epa9A and the second best binder for Epa1A and Epa6A. In general the novel glycan array analysis revealed a strong binding of β1-4-linked galactosides for Epa1A and its mutant variants, while Epa6A, Epa9A and Epa9A^{CBL2Epa1} show almost no binding to all glycans on the chip. The binding specificities of Epa1A^{CBL2Epa6} and Epa1A^{CBL2Epa9} generally resemble that of Epa1A but are not identical. Thus, their binding behavior could be categorized as novel which is in agreement with the results of the CFG analysis.

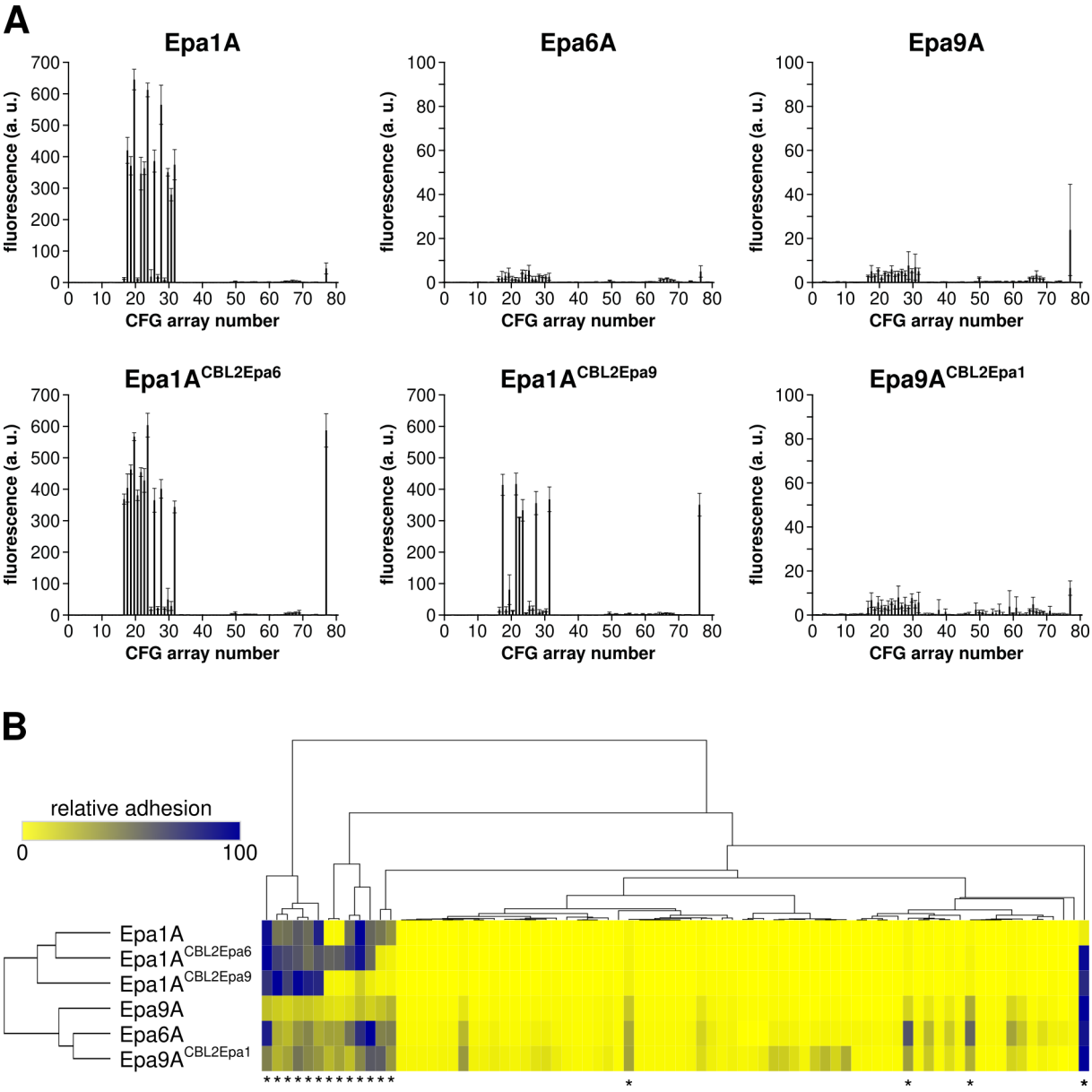

Figure 14: Glycan array data of different EpaA domains binding to novel glycan structures.
A. Three natural EpaA variants, Epa1A, Epa6A and Epa9A, and three chimeric EpaA variants, Epa1A^{CBL2Epa6}, Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1}, were tested on a glycan array chip containing 80 novel complex glycan structures. The adhesion domains were heterologously produced in *E. coli*, purified by a combination of IMAC and SEC, and finally labelled with Alexa Fluor 488 TFP ester as described for the L1 exchange variants in Figure 11. Ligand binding, represented by fluorescence units, is shown by black bars. Error bars indicate the standard deviation. *B.* Clustering of the glycan array results was performed with the Morpheus tool using the Pearson method, that has also been applied for the data shown in Figure 12. Glycans with a terminal Galβ1-4 linkage (*) were bound preferentially by all EpaA variants. A strong binding was found for Epa1A, Epa1A^{CBL2Epa6} and Epa1A^{CBL2Epa9} and an about 100 times lower binding strength was detected for Epa6A, Epa9A and Epa9A^{CBL2Epa1}. A list of all glycan structures present on the chip is shown in Table A3.

2.2.3 Fluorescence titration spectroscopy of EpaA variants

In this work the ligand binding specificities of five different EpaA L1 exchange variants have been characterized in detail by glycan array analysis. However, the semi-quantitative array data do not allow a calculation of binding constants. To address this issue, fluorescence titration spectroscopy

(FTS) experiments were performed to determine dissociation constants (K_D) for natural and chimeric EpaA domains. The T-Antigen (Galβ1-3GalNAc) was chosen as the ligand because Epa1A shows a strong binding to T-Antigen with dissociation constants in the low micromolar range (Maestre-Reyna et al. 2012). Here, dissociation constants were determined for a total of 13 different EpaA variants including the four natural variants Epa1A, Epa6A, Epa9A and Epa10A, the five L1 exchange variants Epa1A^{L1Epa9}, Epa6A^{L1Epa9}, Epa9A^{L1Epa1}, Epa9A^{L1Epa10} and Epa10A^{L1Epa9}, the two CBL2 exchange variants Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} and the two double exchange variants Epa1A$^{CBL2+L1Epa9}$ and Epa9A$^{CBL2+L1Epa1}$ (Figure 15).

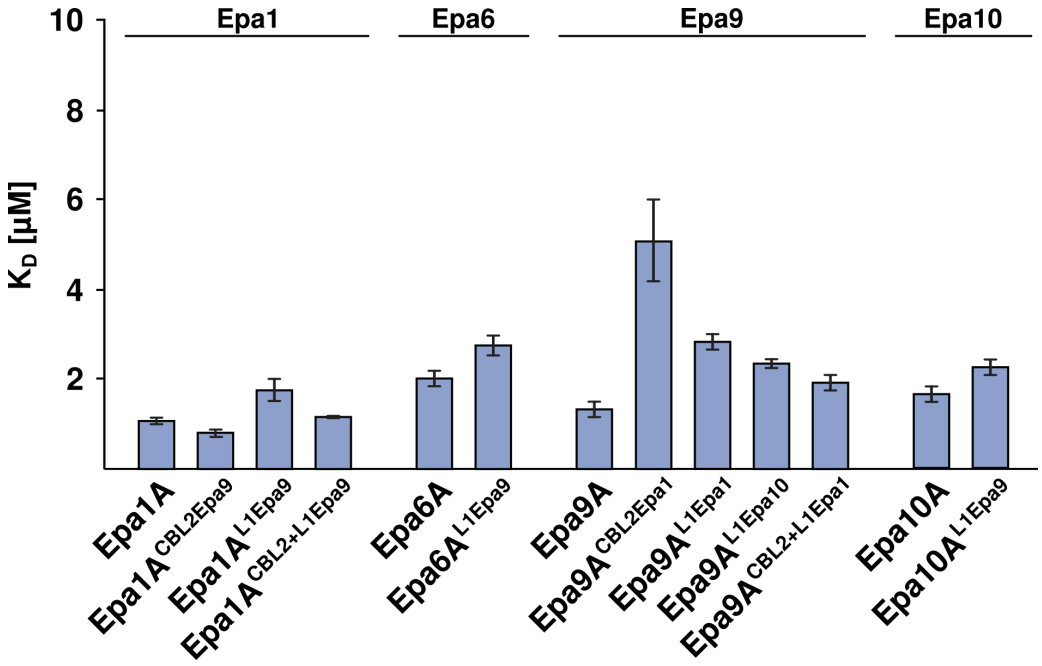

Figure 15: Fluorescence titration spectroscopy of different EpaA variants.
All EpaA variants that were heterologously produced and purified in this thesis were tested for carbohydrate binding to T-antigen by fluorescence titration spectroscopy. In addition, their corresponding natural variants and Epa1A^{CBL2Epa9} were tested as references. Dissociation constants were then determined by nonlinear fitting of titration data (Figure A1). Error bars indicate the standard deviation of three measurements.

For Epa1A a K_D of 1.1 µM was found which is consistent with previous studies (Maestre-Reyna et al. 2012; Diderrich et al. 2015) that reported values of 0.9 µM and 1.7 µM, respectively. Generally, the two CBL2 exchange variants Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} showed a behavior that was similar to the results of the *in vivo* adhesion assay with Epa1A^{CBL2Epa9} being similar to Epa1A and Epa9A^{CBL2Epa1} showing a reduction of T-antigen binding by a factor of five. In the case of the L1 exchange variants Epa1A^{L1Epa9}, Epa6A^{L1Epa9}, Epa9A^{L1Epa1}, Epa9A^{L1Epa10} and Epa10A^{L1Epa9}, the calculated K_D increased in comparison to their respective natural variants, also revealing a reduced ligand binding. Here, the strongest effect was a found for Epa9A^{L1Epa1}, which showed a reduction of K_D by a factor of two. This reduced binding of the L1 exchange variants is in contrast to the results of the *in vivo* adhesion assay where Epa9A^{L1Epa1} and Epa10A^{L1Epa1} showed an increased adhesion. Interestingly, the double exchange mutant Epa1A$^{CBL2+L1Epa9}$ showed an T-antigen binding that was comparable to the natural variant, which also contradicts the results

of the host cell adhesion assay. For Epa9A$^{CBL2+L1Epa1}$, the titration analysis revealed a slightly decreased T-antigen binding, indicated by an increase of the K_D from 1.4 to 1.9 µM in comparison to Epa9A. In summary, the FTS analysis showed an overall reduction of ligand binding for all chimeric EpaA variants. This is in contrast to the results of the *in vivo* adhesion assay which revealed an increased host cell binding for the two variants Epa9A^{L1Epa1} and Epa10A^{L1Epa1}.

2.2.4 Crystallization and structural analysis of CBL2 exchange variants

The CBL2 motif in the inner binding pocket of EpaA domains is a crucial factor for the determination of ligand binding specificity as shown by terminal disaccharide analysis of glycan array data (Figure 13). However, the glycan array data also revealed that an exchange of CBL2 regions between two EpaA variants in most cases was not sufficient to fully transfer the ligand binding pattern. To better understand the relationship between ligand binding specificity and CBL2 structure, the two CBL2 exchange variants Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} were chosen for structural analysis. Both variants were compared with their two corresponding natural variants Epa1A and Epa9A which have already been described in detail in previous studies (Ielasi et al. 2012; Diderrich 2014).

An initial crystallization screen with Epa1A^{CBL2Epa9} in the presence of lactose was done by Lisa Ludewig within the scope of a master thesis (Ludewig 2013). By this approach, crystal growth could be detected within the first 24 hours in over 20 different conditions of the JCSG Core I Suite screen (*Qiagen*). Based on these results, successive screening experiments were performed to optimize the crystallization conditions. This usually leads to the growth of monoclinic crystals of a sufficient size which are suited for structural characterization by X-ray diffraction experiments. Here, growth of lactose (Galβ1-4Glc) containing co-crystals was observed in a condition with 0.1 M MES pH 6.5, 0.2 M ammonium sulfate and 25 % PEG 5000 MME. These crystals were subsequently used for structural characterization experiments which were performed at the BESSY II synchrotron (Berlin, Germany). For Epa9A^{CBL2Epa1}, crystals were found in conditions #57 and #96 of the Classics screening suite and conditions #47 and #70 of screen JCSG Core III. These crystals were characterized by diffraction experiments at the European Synchrotron Radiation Facility (ESRF, Grenoble). For two crystals taken from different conditions sufficient diffraction could be observed (Table 4). The crystal structures for the two CBL2 exchange variants Epa9A^{CBL2Epa1} and Epa1A^{CBL2Epa9} were solved by Viktoria Reithofer from the workgroup of Prof. Dr. Lars-Oliver Essen (Department of Chemistry, Philipps-Universität Marburg) by using molecular replacement. For Epa1A^{CBL2Epa9}, a trimmed model of Epa1A (PDB 4AF9) was used in which the variable loops L1, L2 and L3 as well as the termini regions were removed. The structure of Epa9A^{CBL2Epa1} was solved by using Epa9A (PDB 4CP0) as a template. Refinement statistics for both datasets are given in Table 5.

Table 4: Data collection statistics for Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} complexes

	Epa1A^{CBL2Epa9}·Galβ1-4Glc	**Epa9A^{CBL2Epa1}·Galβ1-4Glc**
PDB code	6Y9J	6Y98
X-ray source	BESSY II BL14.1	ESRF ID23-1
Detector	Dectris Pilatus 6M	Dectris Pilatus 6M
Wavelength (Å)	0.918 41	0.972 42
Space group	$C222_1$	$P3_121$
Cell dimensions (Å)	a = 74.2, b = 104.1, c = 69.1	a = b =65.3, c = 121.9
Resolution range (Å)	34.54 to 1.10 (1.139 to 1.10)	40.63 to 2.80 (2.9 to 2.80)
Observed reflections	212 713	15 594
R_{merge} (%)	2.1 (73.9)	1.8 (44.0)
$I/\sigma(I)$	13.2 (1.0)	19.5 (1.7)
Completeness (%)	99.1 (92.2)	99.2 (100)
Multiplicity	2.0 (1.9)	2.0 (2.0)

Table 5: Refinement statistics for Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} complexes

	Epa1A^{CBL2Epa9}·Galβ1-4Glc	**Epa9A^{CBL2Epa1}·Galβ1-4Glc**
R_{work} (%)	14.6	20.4
R_{free} (%)	16.7	24.8
No. of atoms	2365	1861
Average B factor (Å^2)	19.91	123.13
rmsd bond length (Å)	9	5
rmsd bond angles (°)	1.14	1.04
Ramachandran plot (%)		
Favoured	96.5	93.7
Allowed	3.5	5.0
Outliers	0	1.4

In a next step, the lactose-containing crystal structures of Epa9A (PDB 4CP0), Epa1A^{CBL2Epa9} (PDB 6Y98) and Epa9A^{CBL2Epa1} (PDB 6Y9J) were compared to investigate how the structure of the CBL2 motif determines ligand binding specificity. This comparison reveals that the overall similarity of the three binding pockets is very high. Furthermore, all three crystal structures show an almost identical orientation of the terminal galactose (Figure 16). Despite these structural similarities, the orientation of the secondary glucose moiety differs in all three cases.

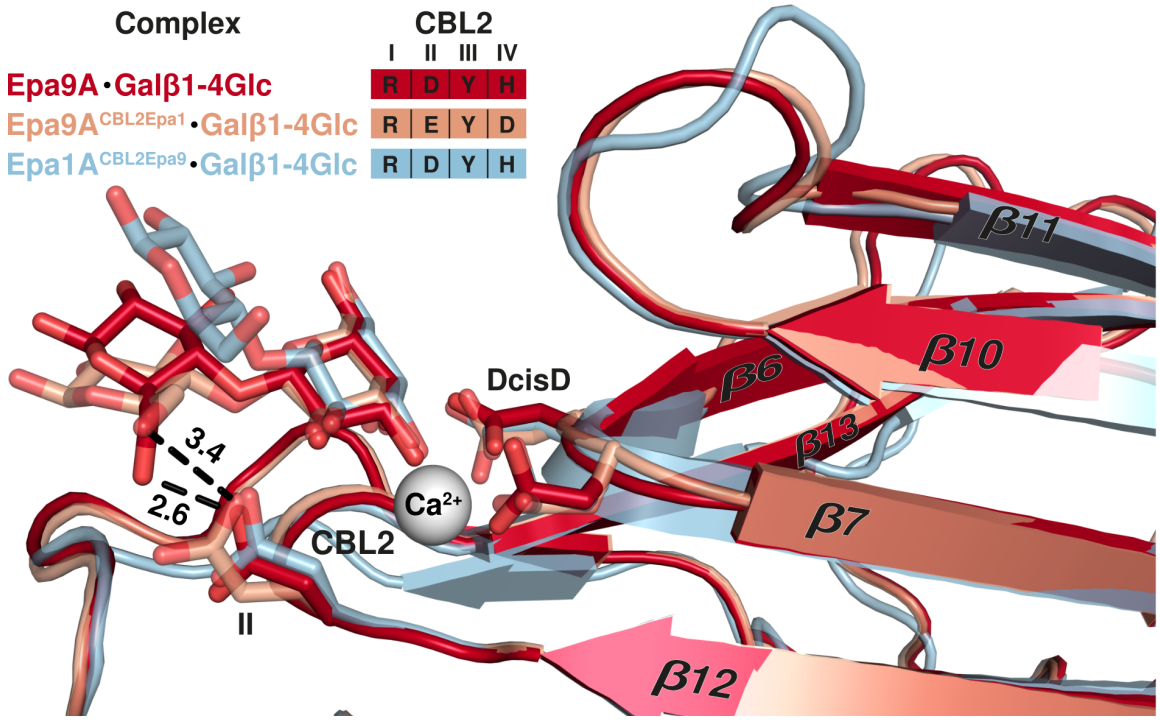

Figure 16: Structural comparison of lactose binding by Epa9A, Epa9A^{CBL2Epa1} and Epa1A^{CBL2Epa9}. The binding pockets of Epa9A (red), Epa9A^{CBL2Epa1} (light red) and Epa1A^{CBL2Epa9} (blue) show an overall high structural similarity. In all three structures the terminal galactose moiety of the lactose ligand shows an almost identical orientation, while the orientation of the secondary glucose differs in each case. For Epa9A^{CBL2Epa1} (PDB 6Y98) the distance (dashed lines) between the 6-OH of glucose and position II of CBL2 is increased in comparison to Epa9A (PDB 4CP0). The crystal structure of Epa1A (PDB 4A3X) was not included in this comparison because it does not contain structural information about the position of the secondary carbohydrate, because this region is not defined by electron density. Corresponding CBL2 sequences for each complex are indicated above.

In the crystal structure of Epa9A, the 6-OH group of the secondary glucose shows interaction with the aspartate residue at CBL2 position II (D258) by formation of a hydrogen bond. This interaction can no longer be observed in Epa9A^{CBL2Epa1} where D258 was replaced with a sterically less demanding glutamate residue. Instead, the secondary glucose shows a different orientation in Epa9A^{CBL2Epa1} and an increase in the distance between the 6-OH and position II of CBL2 from 2.6 Å to 3.4 Å. This structural difference is in agreement with the glycan array data shown in Figure 13 and Figure 14B and indicates an inefficient binding of β1-4 linked galactosides. In the case of Epa1A^{CBL2Epa9}, a direct comparison with Epa1A does not reveal any information on the position of the secondary glucose moiety, since the position of this hexose was not defined by electron density in the available crystal structure (Ielasi et al. 2012). However, in the Epa1A^{CBL2Epa9} complex, the position of the penultimate glucose moiety is defined by electron density, but shows no interaction with the CBL2 motif. Instead, a π interaction between glucose and a histidine of the affinity tag is found which lead to a sufficiently high electron density for the glucose moiety (Figure 17).

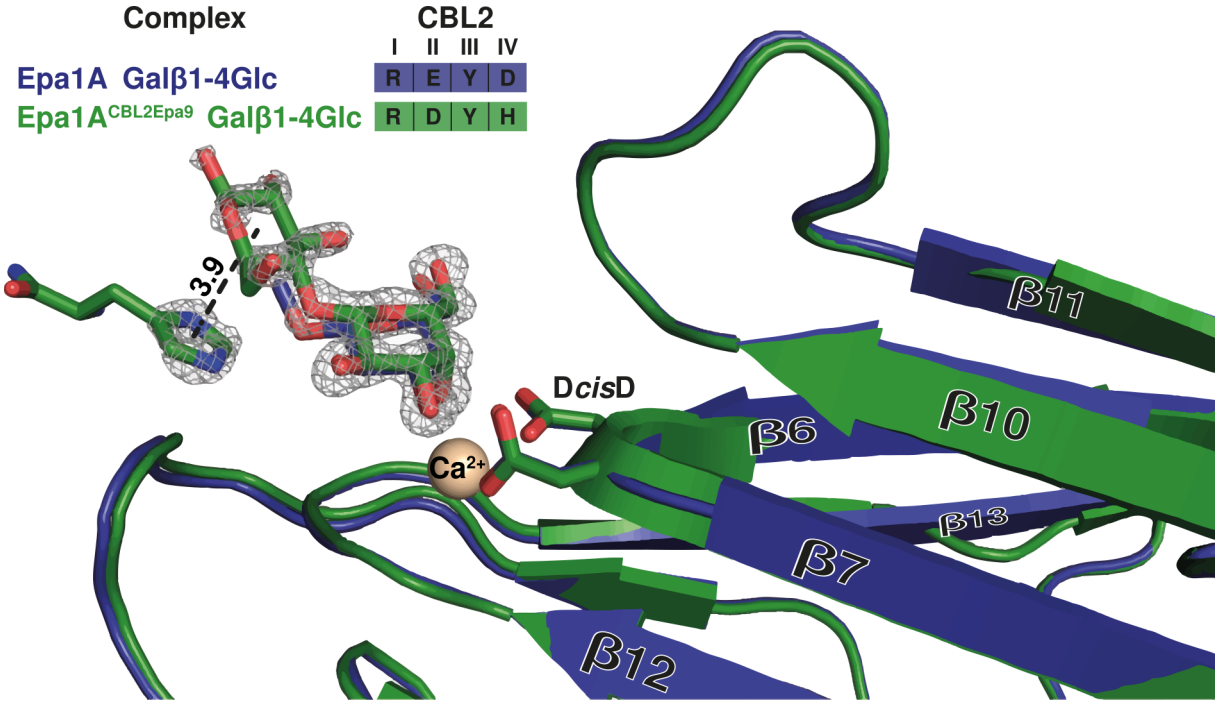

Figure 17: Structural comparison of Epa1A and Epa1A^{CBL2Epa9} bound to lactose.
Structural comparison of Epa1A and Epa1A^{CBL2Epa9} in complex with lactose. The 2Fo-Fc electron density map of lactose (contoured at 2.0 σ) and the interacting histidine of a neighboring epitope tag are shown. For Epa1A^{CBL2Epa9}, the distance between the secondary glucose and the histidine is indicated by dashed lines. CBL2 sequences of the two Epa1A variants are indicated above.

2.2.5 Soaking of Epa9A with complex carbohydrates

The Epa9A crystal structure shows a high overall structural similarity to the adhesion domains Epa1 and Epa6, but differs significantly in the length of the L1 region in the outer binding pocket. Glycan array analysis of all natural EpaA variants of *C. glabrata* strain CBS138 revealed a preferential binding of branched carbohydrates for Epa9A. This specific binding to more complex sugars was suggested to be caused by the extended L1 loop in the outer binding pocket of Epa9A (Diderrich 2014). Since the L1 region of the available Epa9A crystal structures is not defined by electron density this study aimed to provide a co-crystal structure of Epa9A bound to a bi-antennary galactoside comprised of nine sugar moieties which was provided by the workgroup of Prof. Dr. Carlo Unverzagt (University of Bayreuth). Initial crystallization screens were performed with purified Epa9A bound to lactose and the commercially available screens JCSG Core I and JCSG Core II (Qiagen). Crystalline material could be obtained from several conditions of both screens. The three conditions JCSG Core I B10 and JCSG Core II E4 and E5 were chosen for refinement screening to optimize the resulting crystals. Here, several conditions contained suitable crystals which were picked and tested for diffraction at the ESRF beamline 23-1 resulting in two conditions suitable for soaking. Crystals of these two conditions were picked, transferred to the soaking solution and incubated for 0.5 h, 1 h, 2 h, 4 h and 6 h. During incubation, most crystals started to dissolve which resulted in a reduced size of the recovered crystals. Finally, only five crystals could be recovered from the soaking solution. Recovered

crystals were supplemented with 20 % glycerol as cryoprotectant and shock frosted in liquid nitrogen. They were tested at the ESRF beamline 23-1 but unfortunately, no diffraction could be detected for any of the analyzed crystals.

2.3 Molecular dynamics simulation of Epa9A complexes

To date, it was not possible to resolve the structure of the hypervariable L1 region of Epa9A. Previous studies suggested that the elongated L1 of Epa9A could provide additional binding sites for ligand coordination (Diderrich 2014). Following this hypothesis, an improved binding of long and complex carbohydrates would be plausible since in Epa9A the L1 region is extended by a factor of five when compared with Epa1A. In this work, it could be shown that EpaA domains carrying extended L1 regions show reduced *in vivo* adhesion (Figure 10), which contradicts the above hypothesis. At the same time, however, *in vitro* studies with purified adhesion domains showed a strong binding of solute or surface-bound carbohydrates without obvious dependence on L1 length (Figures 13 and 15). To further investigate these apparently contradicting results, the temporal and spatial dynamics of ligand binding was investigated by molecular dynamics (MD) simulations. These computational studies were aimed at providing a model for the L1-ligand interactions and should provide insights into conformational changes of the long L1 region over time. In a first step, an Epa9A model was created with the MODELLER software (Yang et al. 2012) and an Epa9A·lactose complex (PDB 4CP0) as a template. This step was necessary to create a model of Epa9A containing the complete L1 region because the available crystal structures of Epa9A contain no structural information for this loop. Next, three MD simulations were performed to investigate the behavior of the extended L1 loop (i) without a ligand, (ii) with lactose and (iii) with an tetrameric galactoside (gangliotetraose). This oligosaccharide (Galβ1-3GalNAcβ1-4Galβ1-4Glc) has been identified by glycan array experiments to be one of the best bound carbohydrates of Epa9A (Figure 13). All three simulations were performed using the CHARMM force field in combination with the GROMACS software package. The results were evaluated for changes in L1 conformation, protein-ligand interaction and ligand orientation within the binding pocket. Furthermore, the number of putative hydrogen bonds and the protein-ligand interaction energy was calculated for the duration of the simulation.

In a first step, the overall flexibility of the three protein models was determined. Therefore, the root mean square fluctuation (rmsf) of atom coordinates in the protein backbone was calculated. Here, the use of rmsf as a measure of flexibility has an advantages over the use of the root mean square deviation (rmsd) because the rmsf provides a measure for discrete dynamics of individual protein regions rather than an average value for the complete structure. The rmsf for the three trajectories is shown in Figure 18 as a function of residue numbers.

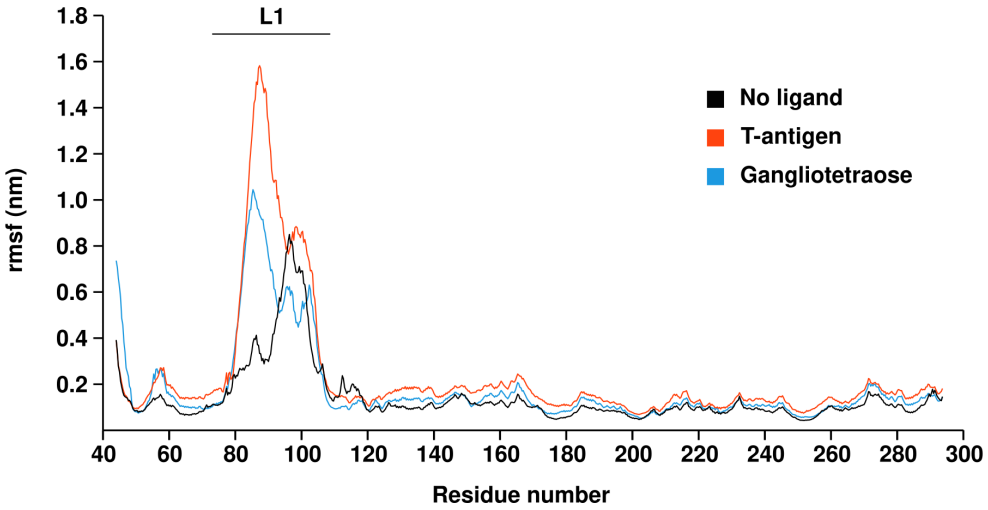

Figure 18: Root mean square fluctuation of Epa9A complexes.
Protein flexibility was determined by calculating the root mean square fluctuation (rmsf) of atoms in the protein backbone. The calculations showed a very low rmsf for the whole length of the Epa9A domain with the exception of the L1 region. Flexibility of the L1 region shows a dependence on the ligand present in the binding pocket. Without a ligand (black) the L1 had a lower rmsf than with T-antigen (red) or gangliotetraose (blue).

In all three cases, the rmsf was generally low with values between 0.1 nm and 0.2 nm but for the L1 region the calculated rmsf was increased by a factor of eight to ten. Interestingly, the flexibility of the L1 region was lower without a ligand (black) than in the presence of T-antigen or the other carbohydrate, showing an rmsf of 0.9 nm. For the models containing T-antigen (red) or the oligosaccharide (blue) maximum fluctuations of 1.6 nm and 1.0 nm were found, respectively. Another interesting observation is that without a ligand especially the mobility of residues 80 to 90 seemed to be limited while in the presence of a ligand these residues show the highest fluctuation in atom position. Generally, this simulations show that the long L1 region of Epa9A is highly flexible which seems to be dependent on the presence of a ligand in the binding pocket.

In addition to calculating the flexibility of the L1 region, the trajectories were evaluated for contacts between protein and ligand and between the L1 loop and the rest of the protein. The first simulation was performed with a model of Epa9A that did not include any ligand. In the first 0.5 ns, the L1 region was in an unordered conformation without contacts to the rest of the protein but then started to orient toward the L2 region in the outer binding pocket Figure 19A. Then, residues D42, T43 and D44 of L1 were in close proximity (2.7 to 3.4 Å) to L2 and stayed in this position until the end of the trajectory (Figure 19B). As a consequence, the overall high flexibility of L1 was reduced and limited to residues D91-Q108, which corresponds to 58 % of the loop. Since the contact between L1 and L2 loops was very stable over a period of 9 ns, the three above mentioned residues were tested for hydrogen bonding with the *FindHBond* tool of UCSF Chimera. This tool uses the distance between possible donor-acceptor pairs and the angle between donor, hydrogen and acceptor to identify possible hydrogen bonds. Criteria for the identification of hydrogen bonds by these two parameters are derived from Mills and Dean (1996). However, no hydrogen bonds could be found between L1 and L2. Since the stable conformation of the L1 region clearly suggested an interaction between both loops, the interaction energy

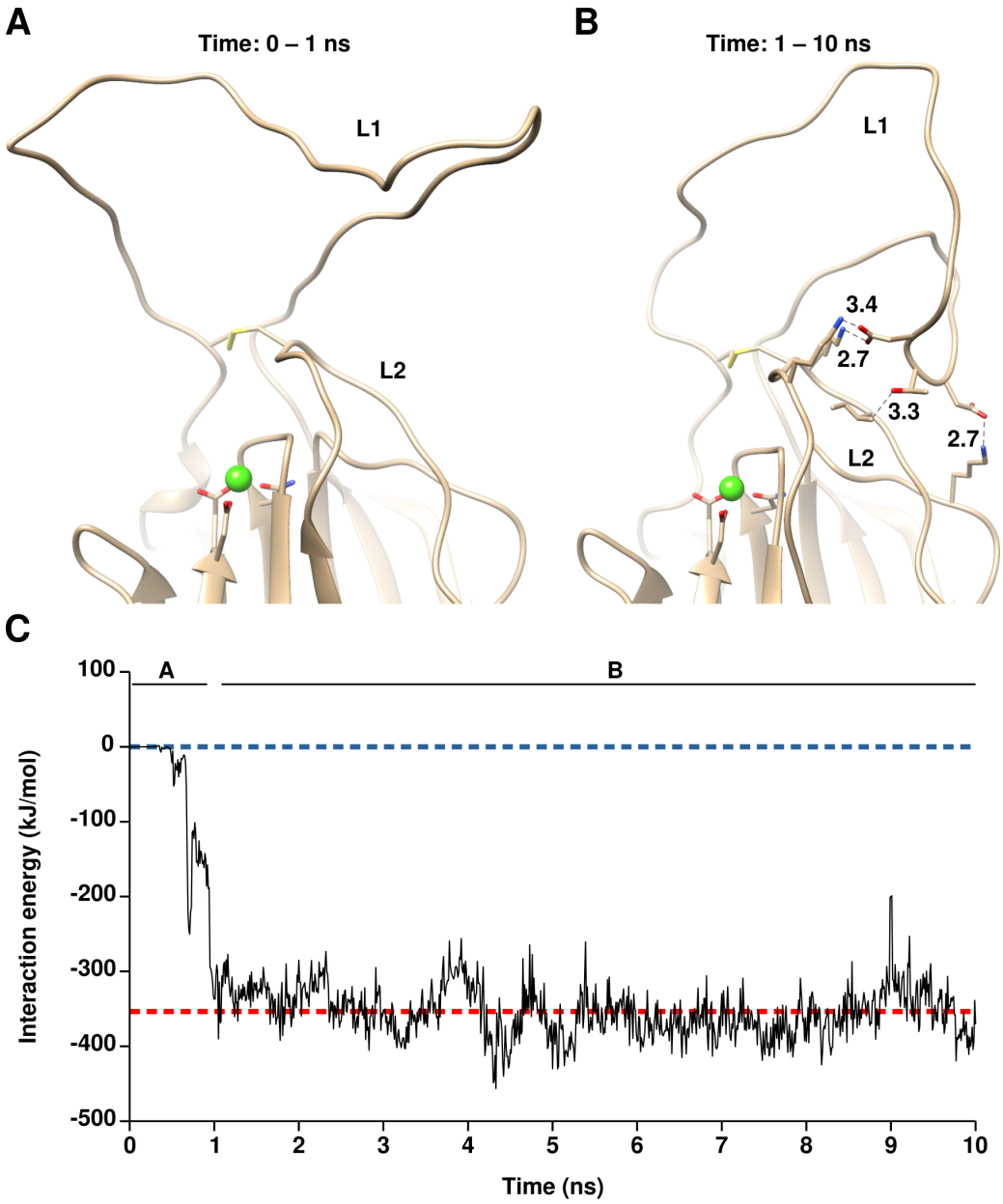

Figure 19: Molecular dynamics simulation of Epa9A showing an interaction between loops L1 and L2 in the absence of a ligand.
The first simulation was performed with a model of Epa9A and a calcium ion (green) in the center of the binding pocket but without a ligand. *A*. At the start of the simulation, L1 was in an unstructured conformation without contact to loop L2. *B*. After 0.5 to 1 ns a contact between L1 and L2 was established that remained stable until the end of the simulation. This interaction involved three residues of L1 (D42, T43 and D44) and four residues of L2 (K105, K109, I111 and K124) with distances between 2.7 to 3.4 Å. *C*. Between the noncontact (blue) and the contact state (red), an interaction energy of about 350 kJ/mol was calculated using the amino acids given in *B*.

between residues D42, T43 and D44 and the L2 was calculated. Therefore, the *energy* tool of the GROMACS simulation software was used, which calculates the interaction energy between charged molecules (Coulomb potential) and between uncharged molecules (Lennard-Jones potential). The combined interaction energy of both terms is shown in Figure 19C. Here, the change between the free, unbound state of the L1 (A) and the coordination toward the L2 (B) is clearly visible as an increase in interaction energy. The difference in interaction energy between the noncontact and the contact state is about 350 kJ/mol which might explain why the connection between both loops remained stable during the simulation.

The second simulation was conducted with an Epa9A model containing T-antigen as a ligand to address the question whether the elongated L1 loop is involved in disaccharide binding. Here, the T-antigen was chosen, because it has been found to be the best bound disaccharide for Epa9A (Diderrich 2014). The analysis of the trajectory revealed no permanent contact between loops L1 and L2 like it was found without a ligand. Instead, the L1 region appeared to be unordered and very flexible over the whole simulation (Figure 20A and B).

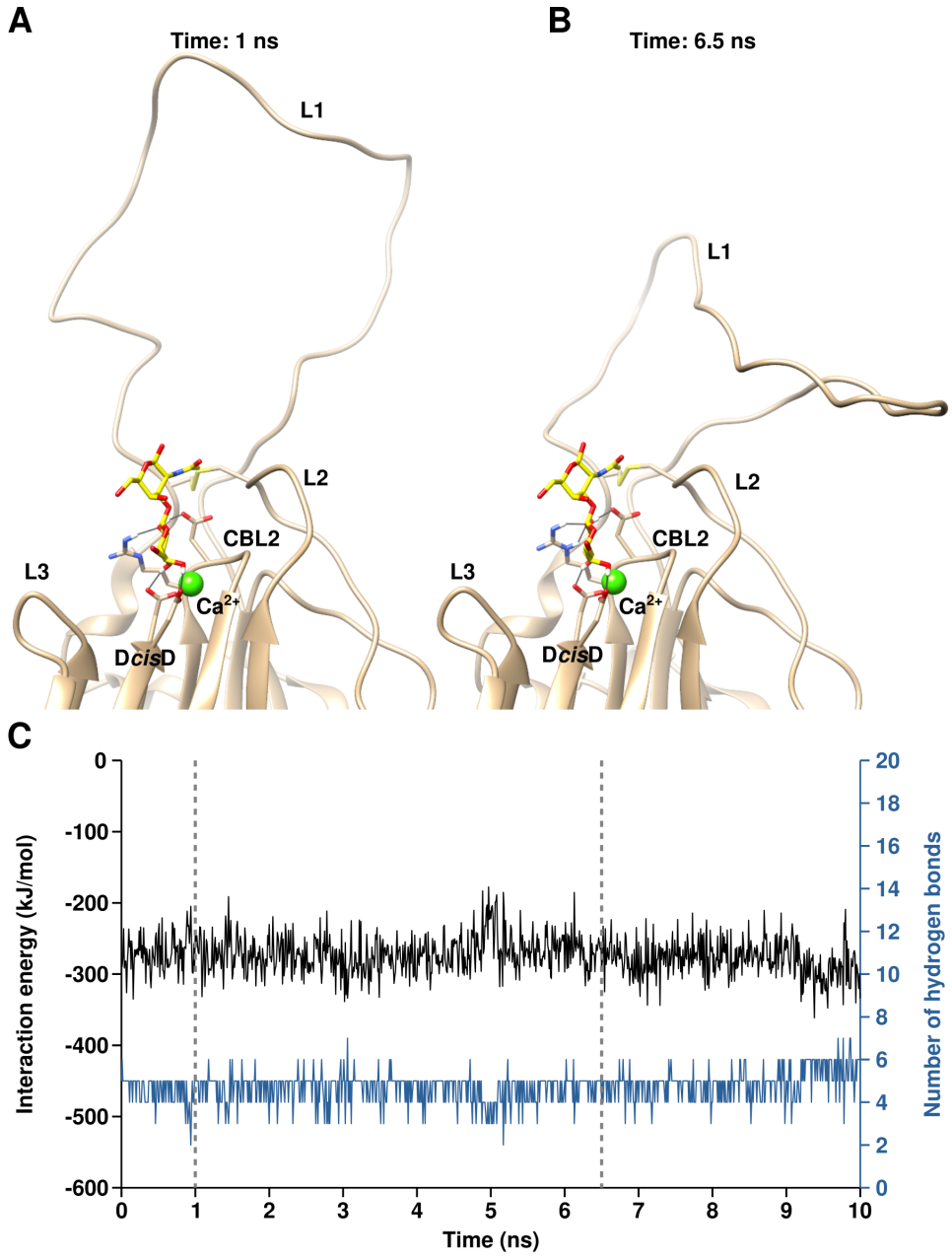

Figure 20: Molecular dynamics simulation of Epa9A in complex with T-antigen.
Protein-ligand interaction between a calcium (green) containing Epa9A model and T-antigen was simulated over a duration of 10 ns. Interaction energy and the number of potential hydrogen bonds were calculated for the protein-ligand complex. *A*. At 1 ns five potential hydrogen bonds between protein and ligand were found, including two contacts between ligand and loop L1. *B*. After 6.5 ns the overall number of hydrogen bonds did not change but in this frame no contacts between T-antigen and the L1 region were found. *C*. Calculated interaction energy (black) and number of hydrogen bonds (blue) between protein and ligand.

The calculated interaction energy for the Epa9A-T-antigen complex was very stable and did not show a dependence on the L1 conformation (Figure 20C). This stable state is also reflected by the calculated number of hydrogen bonds between Epa9A and the ligand. Here, an average number of five hydrogen bonds were found which mostly involved the D*cis*D motif in CBL1 or amino acid residues in the CBL2 motif. This is in agreement with X-ray crystallization data for other Epa9A complexes (Diderrich 2014), which also showed a number of similar contacts between the T-antigen ligand and structural elements of the inner binding pocket. In summary, no direct interaction between the L1 region of Epa9A and T-antigen was found. Furthermore, conformational changes of the elongated L1 region did not seem to influence the protein-ligand interaction in the inner binding pocket.

A third simulation was carried out using gangliotetraose, an oligosaccharide ligand composed of galactose, *N*-acetyl-D-galactosamine and glucose (Galβ1-3GalNAcβ1-4Galβ1-4Glc). The aim of this simulation was to test, whether the L1 region is directly interacting with long carbohydrate structures which protrude from the inner binding pocket and thus might be stabilized by the elongated L1 region of Epa9A. At the start of the simulation, the ligand was bound by a number of noncovalent interactions in the inner binding pocket, involving the D*cis*D motif and the CBL2 region (Figure 21A). Here, the L1 region was in an unordered conformation without making contact to the ligand. After about 1 ns, the glucose moiety (blue) and the L1 were in close contact, leading to the formation of two hydrogen bonds with either end of the loop (Figure 21B). Additionally, a contact between the penultimate *N*-acetyl-D-galactosamine and W110 in loop L1 could be found after 2.4 ns, when the first contact between the L1 region and the glucose moiety could be observed. The contacts between protein and ligand then were stabilized by a conformational change of the L1 region. The loop was further bent toward the ligand and completely covered the glucose moiety at the reducing end of the glycan structure after 3.9 ns (Figure 21C). In this conformation, the L1 region formed a clamp-like structure and thus effectively arrested the ligand in place. This MD simulations showed that despite the high flexibility of the L1 region found in the other simulations, this state remained stable for more than half of the trajectory before loop L1 reverted back to an unordered conformation. So the third simulation could show that for glycan chains with a length of at least four moieties the elongated L1 region can provide additional binding sites and occupy a stable conformation which effectively fixes the ligand position within the outer binding pocket.

In summary, the evaluation of three different MD simulations revealed that the elongated L1 region of Epa9A can either adopt a stable conformation or remain highly flexible, depending on the presence of a ligand. As already hypothesized by previous studies, the L1 showed specific binding of a long ligand but showed no contact to a disaccharide ligand. Surprisingly, loop L1 was bound to the L2 region in the absence of a ligand, which effectively immobilized it for the duration of the simulation. The *in silico* analysis also showed that in complex with the T-antigen, one of the best bound ligands of Epa9A, the L1 region remains flexible and does establish contacts to the disaccharide ligand.

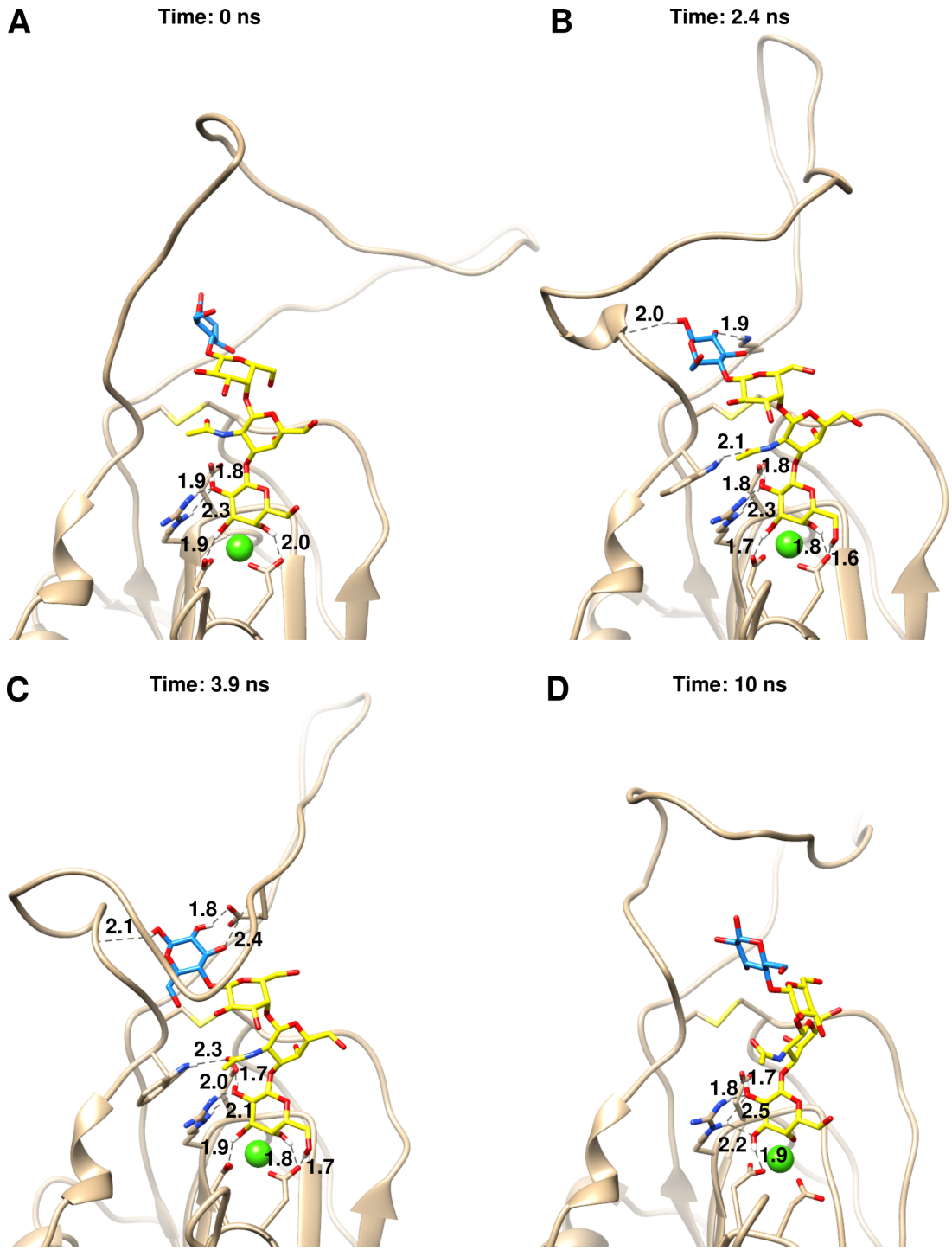

Figure 21: Molecular dynamics simulation of Epa9A in complex with gangliotetraose.
Protein-ligand interaction between a calcium (green) containing Epa9A model and a tetrameric galactoside (Galβ1-3GalNAcβ1-4Galβ1-4Glc) was simulated over a duration of 10 ns. Potential hydrogen bonds and distances were determined with UCSF Chimera. *A*. First frame of the trajectory showing an unstructured L1. The terminal galactose is bound in the inner binding pocket and shows five hydrogen bonds toward the residues forming CBL1 and CBL2. *B*. During the simulation, the distance between L1 and ligand is reduced leading to the formation of two hydrogen bonds involving the glucose residue, one at either side of the L1 region. *C*. Formation of a clamp-like structure by the L1 loop which effectively arrests the ligand in place. *D*. After about 8.5 ns the L1 region unfolds and at the end of the trajectory only five contacts between the inner binding pocket and terminal galactose moiety remain.

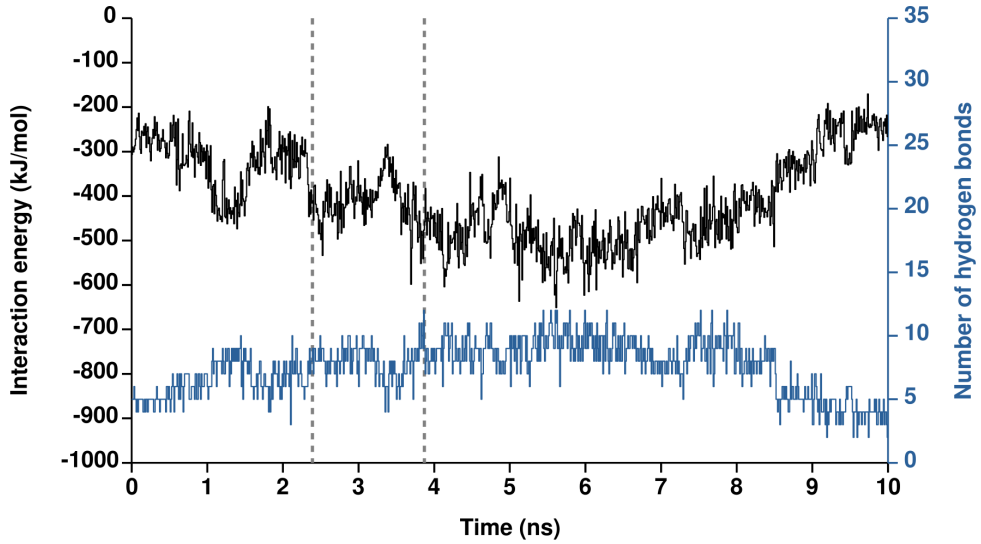

Figure 22: Calculated interaction energy and hydrogen bonds for Epa9A in complex with gangliotetraose. Interaction energy and number of potential hydrogen bonds were calculated for the duration of the trajectory. An increase of interaction energy can be found until about 8 ns. During this time, the L1 approaches the ligand and a number of additional hydrogen bonds are formed. The beginning of L1 folding at 2.4 ns and the maximum overlap between L1 and ligand at 3.9 ns are marked with dashed lines.

2.4 Functional analysis of sulfoglycan-binding EpaA domains

Epithelial adhesins can bind to a wide spectrum of different glycan structures and discriminate between different terminal disaccharides. Among the 17 EpaA domains that have been characterized by glycan array analysis, the four natural Epa adhesion domains Epa12A, Epa15A, Epa22A and Epa23A preferentially bind to sulfated galactosides (Diderrich et al. 2015). Specifically, the best bound glycan for Epa12A, Epa15A and Epa23A carries two sulfate groups, one at C6 of the terminal galactose and one at C6 of the penultimate *N*-acetyl-D-galactosamine ([6S]Galβ1-4-[6S]GlcNAc). Epa22A can bind to a broader range of different glycans but also shows specificity for galactosides with sulfate groups at C6. To this point it is not known which structural features of EpaA domains are responsible for specific binding of sulfoglycans. To address this issue, all four EpaA variants were chosen for creation of mutant variants and *in vivo* and *in vitro* analysis.

A sequence comparison of the four adhesion domains shows that the CBL2 motif differs for all of them in at least one position. The CBL2 motif of Epa12A and Epa15A is nearly identical and just differs in position IV with Epa12A carrying a tyrosine (RDYY), while for Epa15A position IV is an isoleucine (RDYI). In the case of Epa23A, the CBL2 motif differs by two amino acids in positions III and IV (RDFK), while Epa22A carries an unrelated CBL2 sequence (IGKD). The tyrosine at position III of Epa12A and Epa15A was suggested by a previous study to possibly enable an interaction between the hydroxy group of this tyrosine and sulfate moieties (Diderrich 2014). To further address this issue, position IV of Epa12A and Epa15A was mutated to a phenylalanine to remove the polar hydroxyl group. Epa23A already carries an nonpolar phenylalanine at position III, but in this case the lysine at position IV could interact with a sulfate group at C6 of the ligand and was thus exchanged for an alanine. The three resulting variants

Epa12A^{Y227F}, Epa15A^{Y227F} and Epa23A^{K231A} were created by quickchange mutagenesis and subsequently used for adhesion tests on human epithelial cells and FTS experiments. For Epa22A it was not possible to obtain an expression construct with a mutated CBL2 region, in which the lysin at CBL2 position III (IGKD) was changed to an alanine (IGAD).

2.4.1 Host cell adhesion of sulfoglycan-binding EpaA variants

After mutating the CBL2 motifs of Epa12A, Epa15A and Epa23A, the resulting adhesion domains Epa12A^{Y227F}, Epa15A^{Y227F} and Epa23A^{K231A} were functionally characterized. For this purpose, an adhesion test on human colorectal epithelial cells was performed to determine the *in vivo* functionality of the newly created EpaA variants. First, the four natural and three mutant *EPAA* variants were expressed in *S. cerevisiae* strain BY4741 and then analyzed by immunofluorescence microscopy Figure 23A. An empty plasmid and and an expression construct without an A domain were used as controls. Furthermore, Epa1A was included as a reference, since this natural variant shows the strongest adhesion to human epithelial cells. The natural variants Epa12A, Epa15A and Epa22A and the mutated variants Epa12A^{Y227F} and Epa15A^{Y227F} showed fluorescence levels of 49 to 56 %. For Epa23A and Epa23A^{K231A} the fluorescence signals were comparable to that of Epa1A, which was set to 100 %. Despite the different expression levels, all constructs were subsequently tested for *in vivo* adhesion to Caco-2 cells. However, the results of the immunofluorescence microscopy demand a careful evaluation of the adhesion assay.

Determination of host cell adhesion was performed as described for the CBL2 and L1 exchange variants (Figure 10). Here, Epa1A showed a relative adhesion of 40 % (Figure 23B) which is lower than the values obtained in previous adhesion assays. For Epa12A, Epa15A and Epa23A host cell adhesion data have already been published (Diderrich et al. 2015) and thus can be used for direct comparison with the results shown here. Epa12A showed an adhesion strength of 11 % which is comparable to the value obtained in previous studies. In the case of Epa15A the relative host cell adhesion of 28 % was higher than the published value of 20 %. For Epa23A a relative adhesion of 21 % has been reported which is higher than the adhesion of 12 % determined in this study. Lastly, Epa22A was tested for host cell adhesion for the first time and showed an adhesion of 9 %. In comparison to their respective natural variants, the three mutant variants Epa12A^{Y227F}, Epa15A^{Y227F} and Epa23A^{K231A} showed an overall reduced relative adhesion of 8 %, 23 % and 10 %, respectively. Considering the expression levels of each variant, Epa15A^{Y227F} shows a strongly reduced host cell adhesion while for Epa12A^{Y227F} and Epa23A^{K231A} this effect is less pronounced.

In summary, the host cell adhesion assay shows that mutation of either position III or IV of the CBL2 motif shows a reduced host cell binding for Epa12A, Epa15A and Epa23A, respectively. In a next step, it remains to be elucidated whether the reduced adhesion is the result of an impaired sulfoglycan binding.

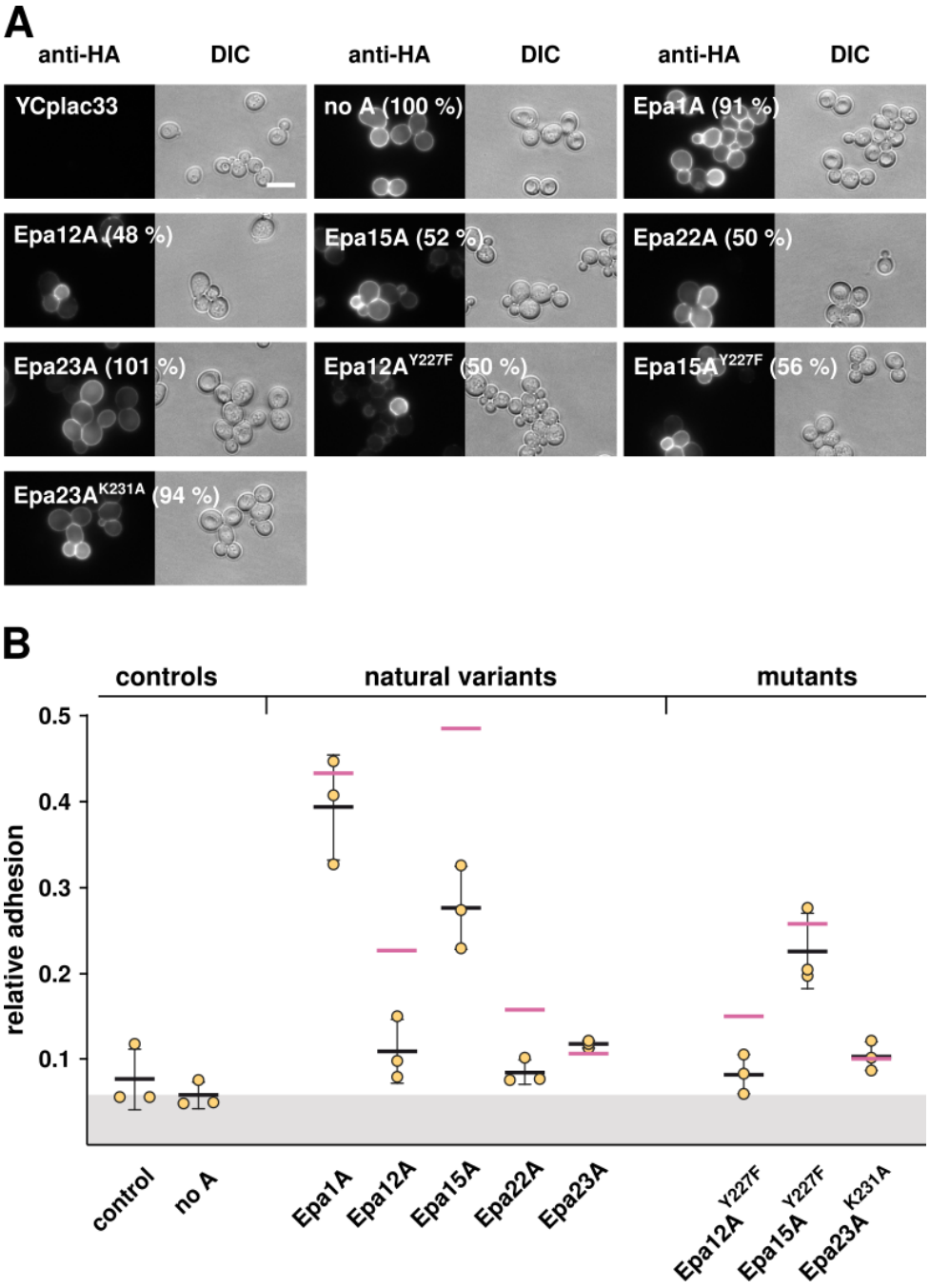

Figure 23: *In vivo* localization of EpaA variants and binding to human epithelial cells.
A. Natural and mutant *EPAA* variants were expressed in *S. cerevisiae* strain BY4741 and tested for proper presentation of the adhesion domains on the cell surface by immunofluorescence microscopy. Labeled cells were visualized and detected by using DIC and a rhodamine filter set (anti-HA). An empty vector and an expression construct without an adhesion domain (no A) were used as controls. Fluorescence intensities of the different variants are shown relative to Epa1A. Scale bar corresponds to 10 µm. *B*. Natural and mutated EpaA variants were tested for adhesion to TCC-SUP cells by using the *S. cerevisiae* strains shown in *A*. Therefore, the yeast strains were grown to logarithmic phase and labeled by Cy3 conjugated antibodies prior to incubation for 2 h on TCC-SUP monolayers. Afterwards, all nonadherent yeast cells were removed by washing with PBS and the amount of adhesive cells was determined by fluorimetry. Relative adhesion was determined by calculating the ratio between adhesive cells and the total amount of cells used for the measurement. Average adhesion (black bars) and standard deviation (error bars) were calculated from three independent measurements (yellow dots). To account for differences in expression shown in *A* the average adhesion was corrected (magenta bars) using the expression rates shown above. Nonspecific adhesion resulting from the carrier domain (no A) is shown by a grey band. The natural Epa1A variant is shown as a reference for strong adhesion. Adhesion assays were performed by Carmen Hütsch in the context of a master thesis.

2.4.2 Production and purification of sulfoglycan-binding EpaA domains

After the characterization of *in vivo* adhesion to human epithelial cells, the mutant EpaA domains were additionally examined *in vitro*. Analogous to the CBL2 and L1 exchange variants, the sulfoglycan-binding EpaA variants carrying mutated CBL2 motifs were produced in *E. coli* and subsequently purified by affinity chromatography and SEC. In addition to the mutant variants the corresponding natural variants were also purified. The purified proteins were then tested by SDS PAGE for purity (Figure 24). Protein bands with a size of about 30 kDa could be found for each protein representing purified protein. These bands correspond to the calculated molecular weight of Epa12A (33.0 kDa), Epa15A (33.0 kDa), Epa22A (29.0 kDa) and Epa23A (33.0 kDa) as well as their corresponding mutants of about the same size. In all three cases additional band were found indicating impurities that could not be separated by either of the chromatography methods. The protein yield of Epa12A and Epa12A^{Y227F} was comparable with 0.5 to 0.8 mg/l. Similarly, both purified variants of Epa15A yielded about 0.1 mg/l of purified protein. For the Epa22 adhesion domain an amount of 0.3 mg/l could be purified while Epa23A and Epa23A^{K231A} yielded 6.7 mg/l and 2.9 mg/l, respectively. In general, all mutant adhesion domains could be produced and purified in a sufficient amount. However, the presence of additional protein bands in the SDS page have to be taken into account in the evaluation of subsequent *in vitro* analyses.

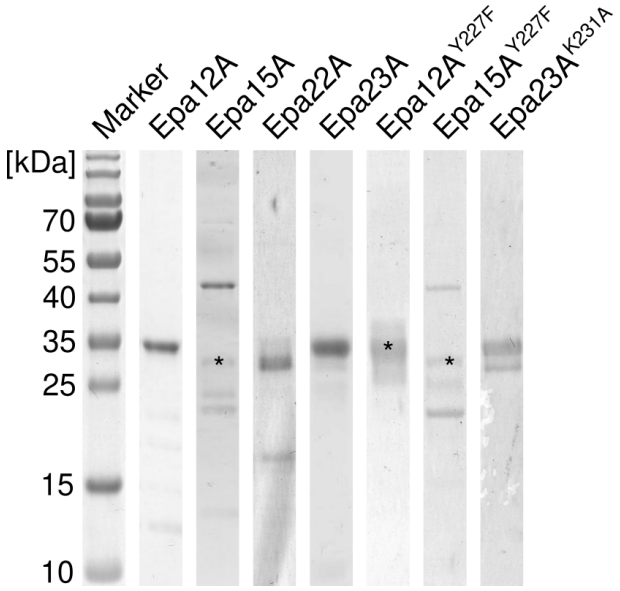

Figure 24: SDS PAGE of mutated sulfoglycan-binding domains after size exclusion chromatography.
Mutant EpaA variants were heterologically produced in *E. coli* and purified by affinity chromatography and size exclusion chromatography. Purity of the eluted protein was tested by SDS-PAGE. In all cases, protein bands of the expected size of 29 to 33 kDa were found. However, in three cases (∗) only a small amount of the target protein with low purity could be obtained. Natural variants were also purified and are shown for comparison.

2.4.3 Fluorescence titration spectroscopy with three different carbohydrates

After production and purification, the newly created EpaA CBL2 variants were characterized *in vitro* by fluorescence titration spectroscopy. Therefore, three different carbohydrates were used. The two monosaccharides galactose (Gal) and 6-sulfogalactose (6S-Gal) were chosen to directly

compare the binding of nonsulfated and sulfated galactosides. Additionally, all EpaA variants were tested for binding of T-antigen. This enables the comparison of mono- and disaccharide binding as well as the direct comparison with already available dissociation constants. Figure 25 shows the calculated dissociation constants for the four natural variants Epa12A, Epa15A, Epa22A and Epa23A as well as the the mutant variants Epa12A^{Y227F}, Epa15A^{Y227F} and Epa23A^{K231A}.

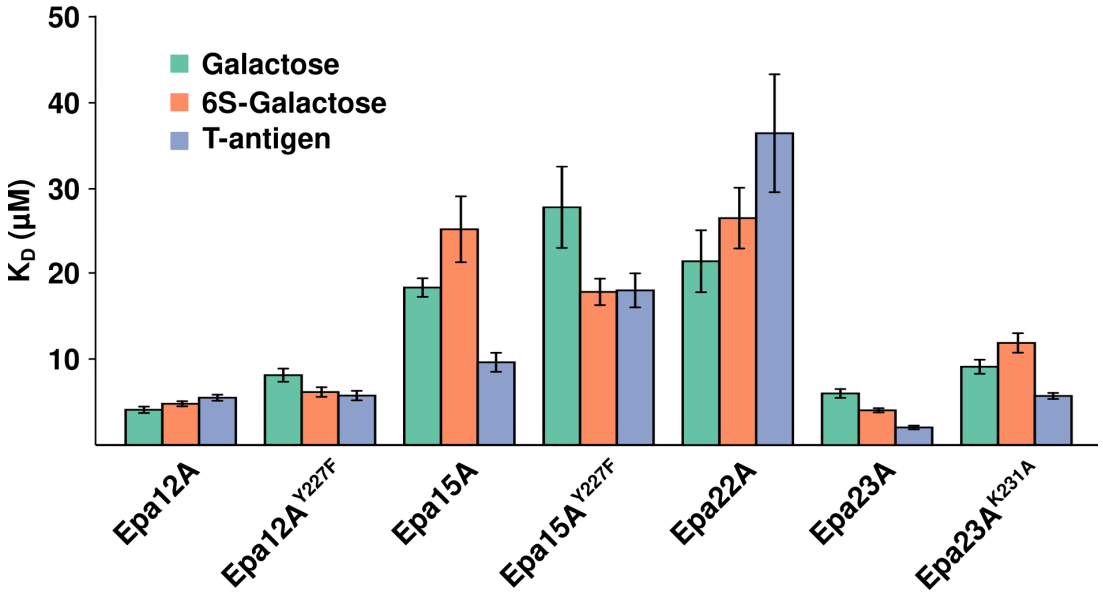

Figure 25: Dissociation constants of sulfoglycan-binding EpaA domains obtained by fluorescence titration spectroscopy.
Carbohydrate binding of the four natural EpaA variants Epa12A, Epa15A, Epa22A and Epa23A was characterized *in vitro* by FTS with galactose (green), 6-sulfogalactose (orange) and T-antigen (blue) as ligands. Additionally, three corresponding variants carrying mutated CBL2 sequences (Epa12A^{Y227F}, Epa15A^{Y227F} and Epa23A^{K231A}) were also tested for binding toward all three carbohydrates. Dissociation constants were determined by nonlinear fitting of titration data (Figures A1, A2 and A3). Error bars indicate the standard deviation of three measurements.

All three mutant EpaA variants showed a weaker galactose binding than the corresponding natural variants. Here, the strongest difference could be found between Epa15A and Epa15A^{Y227F} with dissociation constants of 19 µM and 27 µM, respectively. The adhesion of Epa12A^{Y227F} to 6S-Gal was slightly reduced, while for Epa23A^{K231A} a significantly reduced binding with a K_D of 12 µM was found. In contrast, Epa15A^{Y227F} exhibited a stronger binding of 6S-Gal than the natural variant Epa15A. The adhesion to T-antigen was unaffected in the Epa12A^{Y227F} mutant with a K_D of 6 µM. For Epa15A^{Y227F} and Epa23A^{K231A} the binding of T-antigen was reduced, indicated by an increase in dissociation constants. So far, no data were available for sulfogalactose binding by Epa22A. To address this issue, Epa22A was also characterized by FTS using the three available carbohydrates to allow proper comparison with previous data. Here, the calculated K_D for T-antigen was tenfold higher (35 µM) than reported and binding to Gal and 6S-Gal was quite weak with dissociation constants of 21 µM and 26 µM, respectively. This might be connected to the relatively low yield of the protein purification of 0.25 mg/l which in comparison with other studies is reduced by a factor of 20 and a shows a low purity of the target protein (Figure 24).

In general, the fluorescence titration spectroscopy of Epa12A revealed that the Y227F mutation does not significantly affect the binding of 6-sulfogalactose and T-antigen but leads to a minor reduction of galactose binding. However, in Epa15A the same mutation clearly reduces galactose and T-antigen binding but increases the adhesion to sulfated galactose. This suggests that in Epa12A and Epa15A the tyrosine at position III of the CBL2 motif might be involved in the binding of sulfated glycans but also indicates that additional factors are necessary for proper adjustment of this adhesion. In the case of Epa23A the increase in K_D points to the same conclusion for the lysin at position IV but the K231A mutation caused an overall reduction of binding strength for all tested carbohydrates.

2.4.4 Crystallization of sulfoglycan-binding EpaA domains

In a next step a structural characterization of Epa12A and Epa23A was attempted to investigate how the amino acids at positions III and IV of CBL2 influence the binding of sulfated galactosides. Due to the low yield and insufficient purity of Epa15A no crystallization was performed in this work. For Epa12, a co-crystallization with lactose was performed, as this approach has already been successful with Epa1A, Epa6A and Epa9A. To initially determine suitable conditions for the crystallization of Epa12A a set of commercially available sparse-matrix screens was used. Here, crystal growth could be detected in several different conditions in the “Classics” screening suite and the JCSG Core I screen (Figure 26A-C). As these crystals were too small for characterization by X-ray diffraction, subsequent screens were set up to optimize the crystallization conditions. However, the crystal growth found in the initial screens could not be replicated.

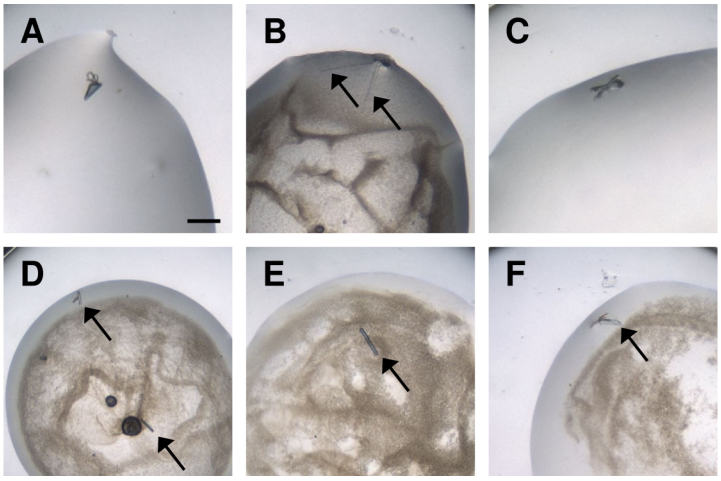

Figure 26: Crystals of Epa12A.
Crystallization experiments for Epa12A were performed with lactose (*A*-*C*) and 6-sulfogalactose (*D*-*F*) as ligands. Lactose co-crystals were found in *A*. condition #42 of Classics screen, *B*. condition #72 of Classics screen and *C*. condition #71 of JCSG Core I screen. In all three cases a protein concentration of 7 mg/ml was used. Co-crystallization with 6S-Gal resulted in crystal growth under three different conditions: *D*. condition #51 of JCSG Core I screen, *E*. condition #71 of JCSG Core I screen with a protein concentration of 10 mg/ml and *F*. Condition #71 of JCSG Core I screen with 5 mg/ml protein. Scale bar corresponds to 100 µm.

In a second approach, lactose was substituted with 6S-Gal in the sparse-matrix screening. This was done to enhance the crystal-formation by providing a more favorable ligand carrying

a sulfate group at C6. Again, the formation of small crystals could be observed in multiple conditions of the JCSG Core I screen using two different protein concentrations (Figure 26D and E). Successive optimization screens based on the three conditions shown in Figure 26D and E yielded a number of crystals that were tested for diffraction. Unfortunately, no diffraction could be detected for any of the tested crystals.

A different approach was used for the crystallization of Epa23A. Here, co-crystallization with 6S-Gal was tested in the initial screens because it is carrying a sulfate group and thus might be favorable for the formation of crystals. This was tested with a total of eight different commercially available sparse matrix screens. A number of small crystals were found in various conditions of the screens JCSG Core I, III and IV while several conditions of "Classics", JCSG+, JCSG Core II, Morpheus and PACT showed spherulite formation as precursors for crystal growth. In order to enhance the crystal growth additional screens were set up based on the conditions shown in Figure 27. Three conditions of these follow-up screens resulted in crystal growth suitable for further characterization. These crystals were picked and analyzed at the ESRF. However, no diffraction could be detected for two of the crystals while one was characterized as salt.

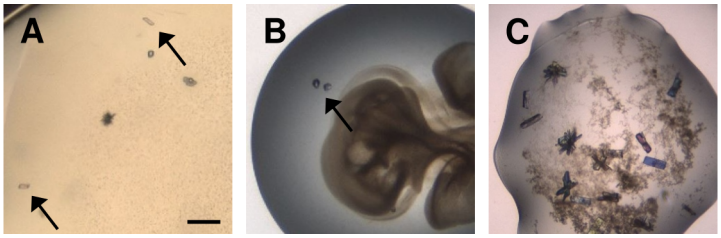

Figure 27: Crystals of Epa23A.
Co-crystallization of Epa23A was done with 6-sulfogalactose and showed crystal growth in three different conditions: *A*. JCSG Core I screen condition #29 with 5 mg/ml protein, *B*. JCSG Core III screen condition #47 with 50 mg/ml protein and *C*. JCSG Core IV screen condition #74 containing 5 mg/ml protein. Scale bar corresponds to 100 µm.

2.4.5 Docking simulations with sulfoglycan-binding EpaA domains

In this work, no crystal structures could be obtained for Epa12A or Epa23A. Thus, these adhesion domains were modeled *in silico* with the software MODELLER. Additionally, Epa15A and Epa22A were also modeled to further investigate the role of CBL2 in sulfoglycan binding. Therefore the available lactose co-crystal structures for Epa1A (PDB 4A3X), Epa6A (PDB 4COU) and Epa9A (PDB 4CP0) were used as templates because all three adhesion domains share sequence identities of more than 40 % (Table 6) and an overall high structural similarity. This structural similarity is illustrated by low rmsd values between Epa1A and Epa6A (0.37 Å over 177 Cα atoms) or Epa1A and Epa9A (0.64 Å over 144 Cα atoms) respectively (Kock 2015). Modeled EpaA domains were subsequently used for computational docking simulations with Autodock Vina (Trott and Olson 2010) and [6S]Galβ1-3[6S]GlcNAc as a ligand (Figure 28). This galactoside was identified by glycan array analysis to be the best bound ligand for Epa12A, Epa15A and Epa23A and among the ten best bound galactosides for Epa22A (Diderrich et al. 2015).

Table 6: Identity of EpaA domains used for modeling

	Epa1A	Epa6A	Epa9A	Epa12A	Epa15A	Epa22A	Epa23A
Epa23A	56	49	39	50	52	44	100
Epa22A	48	46	38	47	47	100	
Epa15A	64	60	43	67	100		
Epa12A	67	55	39	100			
Epa9A	43	44	100				
Epa6A	78	100					
Epa1A	100						

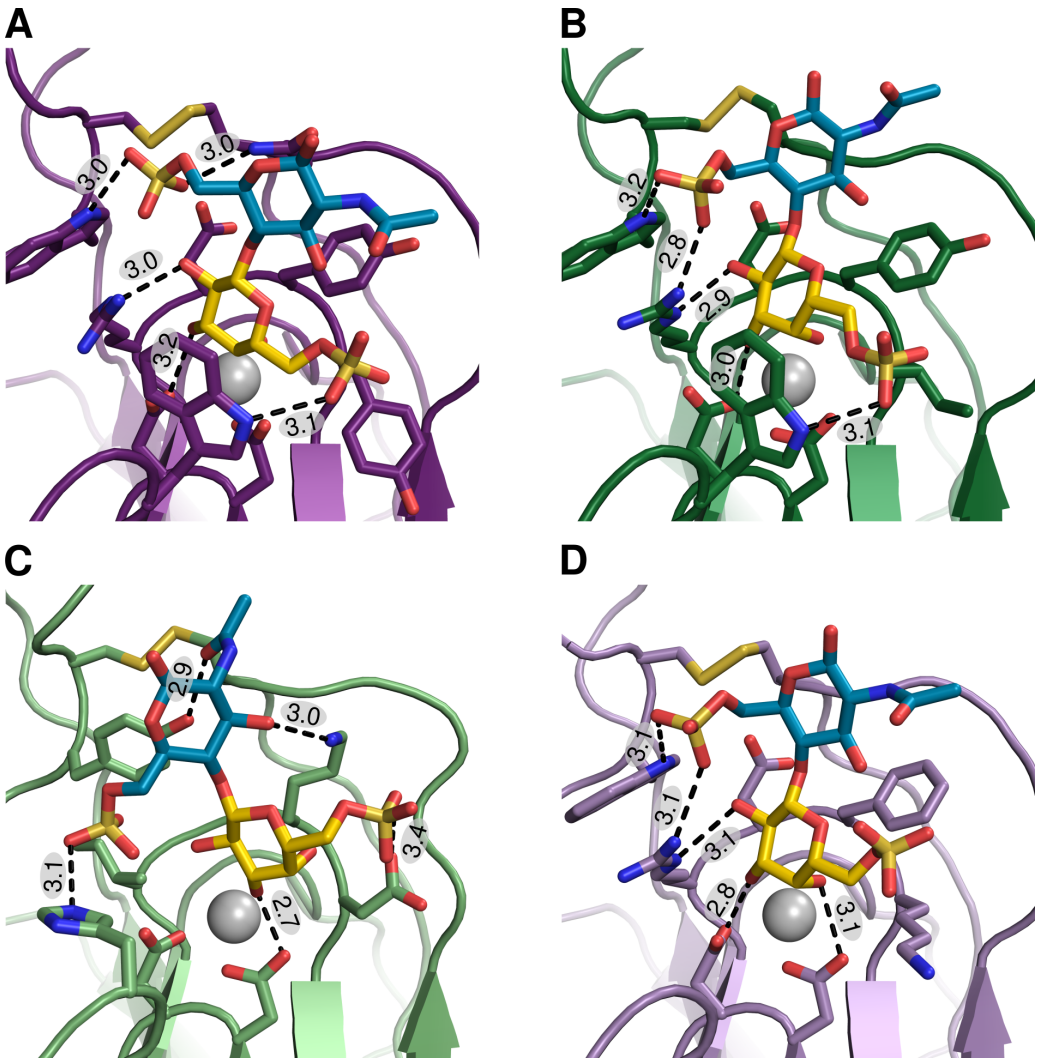

Figure 28: Docking simulation of Epa12A, Epa15A, Epa22A and Epa23A with [6S]Gal**β**1-4[6S]GlcNAc. Molecular docking results for Epa12A (dark purple), Epa15A (dark green), Epa22A (green) and Epa23A (purple) with [6S]Galβ1-3[6S]GlcNAc as a ligand. The calcium ion is shown as a sphere (light green). Putative hydrogen ponds are shown by dashed lines and distances between donor-acceptor pairs are indicated.

The docking simulations resulted in several conformations for the ligand which were first scored according to their free energies of binding. Conformations with the lowest free energies are shown in Figure 28. Potential hydrogen bonds were calculated with the "Show Contacts" plugin of PyMOL and are indicated by dashed lines to illustrate the number of possible interactions between EpaA domain and ligand according to the generated docking model. Terminal sulfogalactose moieties of all four models are coordinating the calcium ions in the center of each binding pocket. This is in agreement with the crystal structures of Epa1A, Epa6A and Epa9A. In contrast to the initial hypothesis, the models of Epa12A and Epa15A show no interaction between the tyrosines at CBL2 positions III and IV and the ligand and thus indicate that these residues are not directly responsible for programming the specificity. The same was found in Epa23A for the lysine at CBL2 position IV. Interestingly, a number of other residues were found that putatively form hydrogen bonds with the ligand. In the models of Epa12A, Epa15A and Epa23A the sulfate group at C6 of the GlcNAc moiety form hydrogen bonds with a tryptophan residue (W37/38) that is located next to the cysteine in L1. A similar interaction can be found for the arginine in the CBL2 of these three models which interacts with the hydroxy groups at C2 or C3 of the [6S]Gal. In Epa12A this arginine also shows a possible interaction with the sulfate group of the glucosamine which suggests that this conformation is probably preferred by carbohydrates with sulfated secondary sugars. Furthermore, Epa12A shows two more stabilizing contacts that involve a glutamine in L2 and the tryptophan in L3.

An interesting finding is that the model of Epa22A, despite its relatively low sequence identity, also includes contacts between the ligand and loops L1 and L3. Here, L1 harbors a tyrosine (Y38) that is involved in ligand binding, but unlike in Epa12A, Epa15A and Epa23A, the tyrosine does not bind to the sulfate group of [6S]GlcNAc, but to the N-acetyl part of the hexose. Instead, the sulfate group in this case is bound by a histidine (H158) in L3. The model of Epa23A does not include the elongated part of the L1 loop because this region could not be modeled due to a lack of structural information for this part of the protein. Overall, the ligand position is very similar in the docking studies with Epa12A, Epa15A and Epa23A. With Epa22A the ligand is tilted which brings the C6 sulfate of the galactosamine moiety closer to L3 and enables an interaction with a histidine in this loop. Positions III and IV of CBL2 did not interact with the ligand with the exception of the Epa22A model, where each of the positions is contacted by one of the ligand moieties, respectively. This computational analysis shows that in most cases conserved residues are responsible for ligand binding, a surprising finding since in other cases the CBL2 region seems to have a considerably bigger impact on ligand binding specificity.

3 Discussion

The adhesion to target cells is a crucial step during an infection of host tissue by many pathogens. Of the 81 adhesin-like CWPs found in *C. glabrata* strain CBS138 (Xu et al. 2020), the large family of epithelial adhesins is best studied and has been shown to be responsible for binding to host glycans. These proteins carry a number of conserved and variable structural elements that enable a high flexibility in ligand binding (Diderrich et al. 2015). However, the complex mechanisms and structural features that are used to define a certain specificity are not yet fully understood. Previous studies have shown that a group of three highly conserved residues located in the binding pockets of EpaA domains are essential for efficient host cell binding (Maestre-Reyna et al. 2012; Diderrich 2014). This group comprises the D*cis*D motif in CBL1, the aromatic side chain of the tryptophan residue in loop L3 and the disulfide bond linking loops L1 and L2.

In this study, a comprehensive structural and functional analysis was performed to analyze how variable structural elements affect ligand binding specificity and host cell adhesion. To address this questions, chimeric EpaA variants have been created in that either the CBL2 motif, the L1 region or both were exchanged with the corresponding elements of other natural EpaA domains. Chimeric adhesion domains were constructed by using the structurally well characterized natural EpaA variants Epa1A, Epa6A and Epa9A. Additionally, Epa10A was included as it differs from Epa9A in ligand binding behavior (Figure 13) although both adhesion domains are closely related which includes an identical CBL2 motif and an L1 region of similar length. The newly created exchange variants were characterized by a number of *in vivo* and *in vitro* experiments to determine their host cell adhesion and ligand binding specificity. Additionally, interactions between the elongated L1 region of Epa9A and different glycan structures were tested by soaking of protein crystals and MD simulations. These were aimed at providing insights about the flexibility of an elongated L1 region in combination with different ligands because to date it was not possible to resolve the structure of this highly variable region.

In a second part of this work four EpaA domains that specifically bind sulfated galactosides were characterized in detail. Here, a mutational analysis of the four adhesion domains Epa12A, Epa15A, Epa22A and Epa23A was performed to address the question how they are programmed to preferably bind to sulfated ligands. Therefore, several *in vivo* and *in vitro* analyses were performed to determine host cell adhesion and ligand binding specificity of the newly created mutants. To also gain information about the structural basis for sulfoglycan binding several crystallization experiments and docking simulations were performed.

3.1 The CBL2 motif is crucial but not sufficient to program ligand binding specificity

In the first part of this study, two variable structural elements of Epa adhesion domains, the CBL2 motif and the L1 region, were analyzed by structure-based mutational analysis. While the

CBL2 motif has been described to be a central factor for the determination of ligand binding specificity and to be crucial for Epa functionality, the L1 region has not been addressed by functional analysis before (Figure 29).

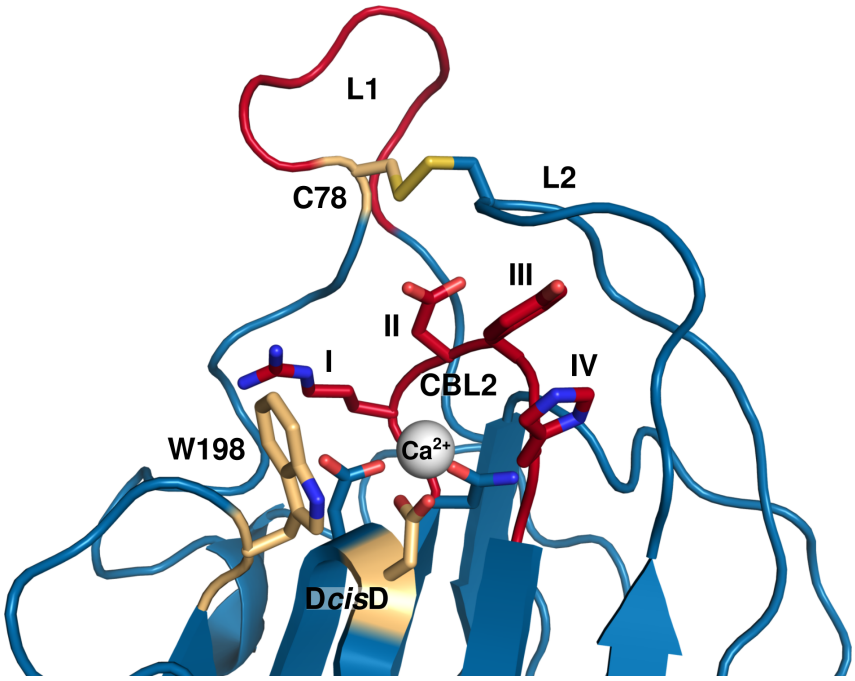

Figure 29: Structural elements of the Epa1A domain that have been analyzed by mutational analysis.
The functionality of conserved and variable elements of the inner and outer binding pocket of Epa1A has been experimentally addressed by structure-based mutational analysis. In previous studies the amino acid residues C78, D165A and W198 (orange) were mutated to alanine and, in the case of W198, additionally to tyrosine and histidine, to determine their role in EpaA-mediated adhesion to human epithelial cells and ligand binding. Here, a number of chimeric EpaA variants were analyzed in which either the CBL2 motif , the L1 region or both (red) were exchanged between the A domains of Epa1, Epa6, Epa9 and Epa10.

Initial glycan array experiments characterized the ligand binding patterns of Epa1A, Epa6A and Epa7A and suggested that the CBL2 motif is involved in ligand binding (Zupancic et al. 2008). This hypothesis is in agreement with other studies that revealed a strong correlation between the sequence of the CBL2 and ligand binding specificity (Maestre-Reyna et al. 2012; Ielasi et al. 2014; Diderrich et al. 2015). The correlation was established by analysis of numerous natural and mutated EpaA variants. These mutant EpaA domains were created either by random mutagenesis of CBL2 positions II and III or the exchange of complete CBL2 motifs between Epa1A, Epa2A, Epa3A, Epa6A and Epa9A. Both approaches showed that in most cases a modification of the CBL2 motif alters host cell adhesion and ligand binding specificity. In this work, the changes in *in vivo* and *in vitro* adhesion have been systematically categorized for the first time to allow a direct comparison between the effects of CBL2 and L1 exchanges. Furthermore, the categorization into recipient, donor and novel binding behavior was used to determine whether a directed reprogramming of EpaA mediated adhesion is possible by exchange of (i) the CBL2 motif, (ii) the L1 region or (iii) both elements. First, the results were evaluated for the available CBL2 exchange variants in comparison with their respective recipient and donor domains. Despite Epa10A being closely related to Epa9A no *in vivo* adhesion data were available

for this natural variant. Thus in this work host cell adhesion was determined for Epa10A and the novel exchange variant Epa10A^{CBL2Epa1} to address this issue and complement the present data. In general, the categorization of CBL2 exchange variants showed an altered host cell adhesion and ligand binding pattern for most chimeras (Table A1). Five of the analyzed mutant variants also exhibited an overall reduced binding strength towards the glycans used in the array. This finding underlines the importance of this variable structural motif for *in vivo* EpaA functionality. Furthermore, ligand binding patterns of most CBL2 exchange variants only partially resemble the recipient and/or donor profiles but retain galactosides as their best binders. These results suggest that the host cell recognition and ligand binding functionality of the EpaA domains in principle can be changed by replacing CBL2 motifs, but that this replacement is not sufficient for a complete reprogramming of EpaA functionality. Consequently, other factors must also be involved in defining functionality. These factors could include a direct ligand interaction by the variable loops L1 or L2 in the outer binding pocket or additional influences like hydration or long-range effects that already have been found in other PA14-like lectins (Veelders et al. 2010).

To further investigate the effects of CBL2 exchanges on ligand binding specificity a novel glycan array designed by the working group of Prof. Dr. Carlo Unverzagt of the University of Bayreuth was used to characterize binding patterns of Epa1A^{CBL2Epa9}, Epa6A^{CBL2Epa9} and Epa9A^{CBL2Epa1} (Figure 14B). The finding that Epa1A strongly binds to β1-4-linked galactosides while Epa9A completely lacks this ability is surprising because in the CFG glycan array analyses the binding of β1-4-linked galactosides was strong for Epa9A and moderate to low for Epa1A (Diderrich et al. 2015). In the case of Epa1A^{CBL2Epa9}, a partial transfer of the weak donor binding behavior to the recipient underscores the view that the CBL2 motif is of central importance for programming the ligand binding specificity. However, a complete reprogramming toward the donor specificity or a novel specificity could not be found, which might be attributed to the smaller number of glycans on this chip compared to the CFG glycan array (Table A2, Table A3). For a glycan array analysis of ligand binding patterns the higher number of glycan structures on the CFG chip allows a more fine-grained analysis of ligand binding. This suggests that the novel chip can be used to complement the available CFG data, but on its own does not provide sufficient data for the analysis of the mutated EpaA domains that are analyzed in this study. Another limitation of the novel array is that only one glycan with a terminally β1-3-linked galactose is present on the chip. This effectively limits the value of the obtained data because many EpaA variants preferably bind to Galβ1-3-linked galactosides (Zupancic et al. 2008; Diderrich et al. 2015). Therefore, it would be useful to add a number of β1-3-linked galactosides to the chip to be able to compare different sets of glycan array data.

In summary, the novel glycan array chip allowed to test the ligand binding specificity of different EpaA variants toward a large number of yet untested glycan structures and thus adds novel information about the binding specificity of these adhesion domains. However, due to the lack of β1-3-linked galactosides on the chip the novel data cannot be directly compared with previous experiments. Nevertheless, the results of both glycan arrays indicate a reprogramming

of binding specificity by exchange of the CBL2 motif and hint toward the influence of additional structural elements because only an incomplete transfer of binding patterns could be detected.

An additional aim of this study was to gain further insights on how variable structural motifs determine the ligand binding specificity of Epa adhesion domains. Initial structural studies revealed that in Epa1A and Epa6A the CBL2 positions II and III are responsible for discrimination between α- and β-linked galactosides in ligand binding (Maestre-Reyna et al. 2012; Diderrich et al. 2015). In the present study, a structural analysis of the two CBL2 exchange variants Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} was performed to produce high resolution structures allowing a direct comparison of both mutants with their respective natural variants. This analysis revealed that the orientation of the CBL2 motif is identical within the lactose co-complexes of Epa9A and Epa1A^{CBL2Epa9}. Simultaneously both variants exhibit significant differences in the binding mode of the secondary glucose moiety (Figure 16). By this observation it seems very likely that other structural elements are also involved in programming ligand binding specificity because here the CBL2 motif cannot be responsible for the observed differences in binding modes. The identical orientation of the CBL2 motifs also explains the missing specificity for α-linked galactosides in Epa1A^{CBL2Epa9} as this feature might not be directly programmed by the CBL2 motif and thus could not be transferred from Epa9A to Epa1A by CBL2 exchange (Figure 13). A similar case has previously been found for Epa7, which shows a significantly better discrimination between α- and β-linked 1-3 galactobiose than Epa1A (Diderrich et al. 2015). Despite this difference in ligand binding, their binding pockets differ only by individual residues in loops L1 and L2, suggesting that these residues may indirectly alter the structure of the Epa7A binding pocket, leading to a more pronounced ligand binding specificity. In order to investigate this possibility for Epa9A and Epa1A^{CBL2Epa9}, a detailed structural analysis of protein-ligand complexes with α-linked galactosides could be performed. However, it was not yet possible to obtain suitable co-crystals of the two variants Epa9A and Epa1A^{CBL2Epa9}.

Another interesting finding was that Epa1A^{CBL2Epa9} exhibits an additional specificity for sulfated galactosides which can also be found in the other CBL2 exchange variants Epa1A^{CBL2Epa2}, Epa1A^{CBL2Epa3} and Epa3A^{CBL2Epa1}. Obviously, the exchange of the CBL2 motif fosters the binding of sulfoglycans, although neither the recipient nor the donor variant shows a binding of sulfated sugars. An initial docking simulation using the crystal structure of Epa1A^{CBL2Epa9} as a template and [6S]Galβ1-3[6S]GlcNAc as a ligand revealed an interaction between the terminal galactose and sulfate groups of the ligand with conserved elements of the binding pockets (Figure 30). However, a proposed binding of sulfated sugars by positions III and IV of CBL2 (Diderrich 2014) could not be found. Based on these data it appears questionable that the exchange of the CBL2 motif has a direct impact on sulfoglycan binding but rather indirectly influences binding toward these ligands. To further investigate how the exchange of the CBL2 motif enables this novel specificity, it would be interesting to analyze the structure of Epa1A^{CBL2Epa9} in complex with sulfated galactosides. Unfortunately, no suitable ligands are currently available, which complicates the further analysis of ligand binding by Epa1A^{CBL2Epa9}.

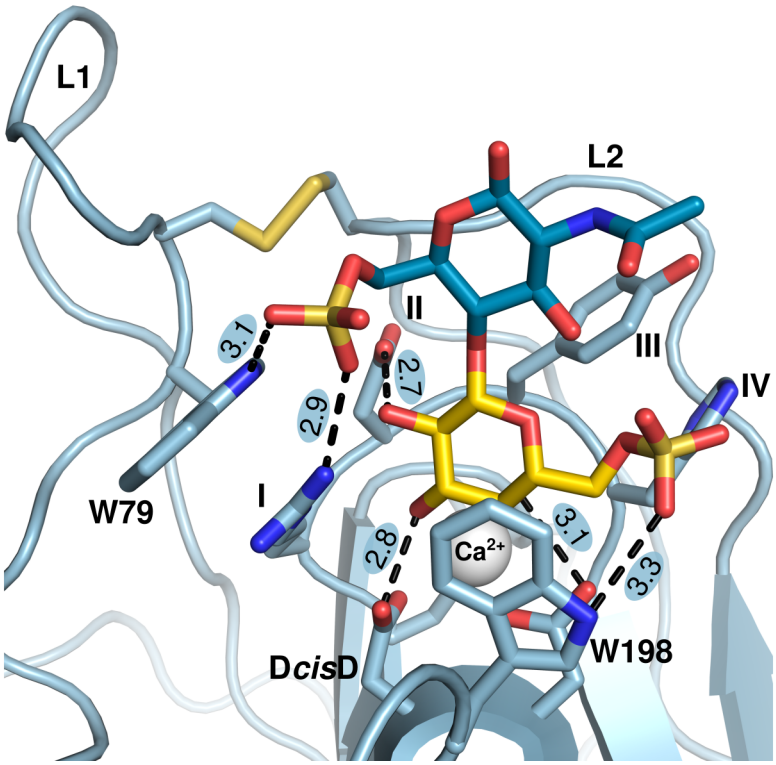

Figure 30: Docking simulation of Epa1A^{CBL2Epa9} bound to [6S]Gal**β**1-3[6S]GlcNAc.
Binding of sulfoglycans by Epa1A^{CBL2Epa9} was investigated structurally by performing docking simulations with [6S]Galβ1-3[6S]GlcNAc as a ligand. The simulation showed possible contacts between the ligand and conserved residues in CBL1, CBL2, L1 and L3 which also included both sulfate groups. Interestingly, neither position III nor position IV of CBL2 were involved in ligand binding which contradicts the assumption that ligand binding specificity is controlled by variable structural elements.

A second comparison using Epa1A and Epa9A^{CBL2Epa1} reveals a very similar spatial orientation of their CBL2 motifs, especially of residues II and III, in both complexes. In addition, both variants have an almost identical ligand binding profile (Figure 13) which in this case suggests that by certain CBL2 exchanges the ligand binding specificity of the donor can be fully transferred toward the recipient. However, Epa9A^{CBL2Epa1} exhibits a reduced binding strength in both host cell adhesion and glycan array analysis (Table A1) which indicates that the overall functionality of this variant might be impaired. The reduced binding strength in host cell binding and glycan array analysis is contradicted by the results of the FTS measurements because in the latter experiments Epa9A^{CBL2Epa1} showed a strong binding to T-antigen in the lower micromolar range. This disagreement could be caused by the use of T-antigen, a disaccharide, in FTS while in glycan arrays and *in vivo* binding studies longer and more complex sugar structures are present. Furthermore, in FTS the disaccharide ligands were in solution, while in host cell adhesion experiments and glycan arrays they were surface attached. Thus, strong T-antigen binding in FTS could possibly be attributed to an alleviated ligand binding resulting from the additional degrees of freedom of the ligand in solution (Du et al. 2016). To address this issues, it could be beneficial to try other methods for K_D determination. Here, surface plasmon resonance or atomic force microscopy could be tried as as both methods have already been used to characterize the interaction between Epa proteins and their ligands (Zajac et al. 2016; Valotteau et al. 2019).

An overlay of the inner binding pockets of Epa1A, Epa9A, Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} reveals that all four share a highly similar structural orientation of the CBL2 with the exception of the aspartate (D260) at position IV of Epa9A^{CBL2Epa1} (Figure 31). Here the side chain of D260 points away from its original orientation in Epa1A and is rotated by nearly 90°.

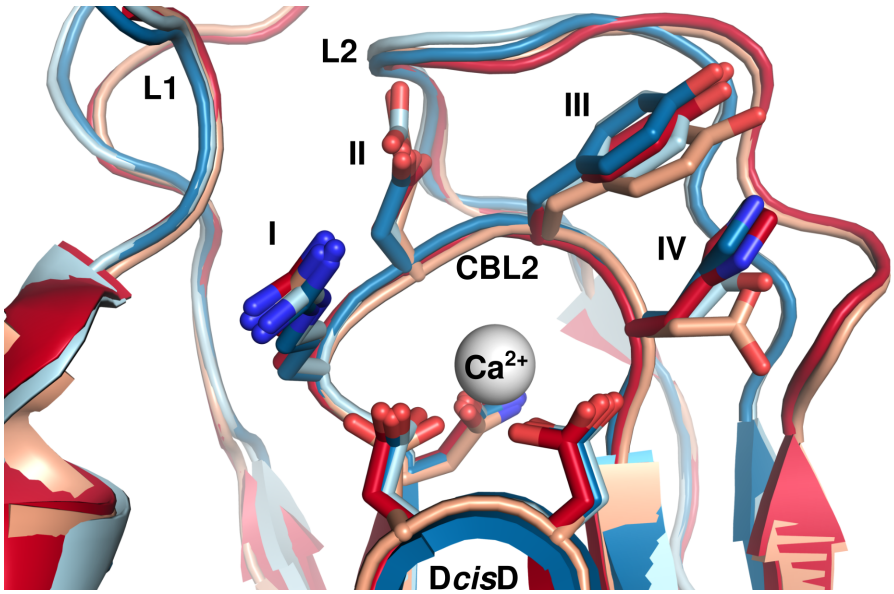

Figure 31: Overlay of Epa1A, Epa9A, Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} binding pockets.
The spatial orientation of amino acids in the CBL2 motifs was analyzed by creating an overlay of Epa1A (light blue), Epa9A (red), Epa1A^{CBL2Epa9} (blue) and Epa9A^{CBL2Epa1} (light red). Therefore, the corresponding crystal structures 4A3X, 4CP0, 6Y9J and 6Y98 were used. The comparison shows a nearly identical structure for the inner binding pockets of all four adhesion domains. One minor difference can be found in the CBL2 motif at position IV, where the aspartate side chain of Epa9A^{CBL2Epa1} is positioned at a different angle than in Epa1A.

However, since in all four structures the residue at CBL2 position IV is not involved in ligand binding this structural difference seems to have no direct influence on the functionality of the adhesion domain. The comparison of Epa1A and Epa1A^{CBL2Epa9} also reveals a highly similar conformation of the binding pockets in both structures. Despite this similarity, both adhesion domains show significant differences in ligand binding specificity, which is illustrated by the improved binding of sulfoglycans by Epa1A^{CBL2Epa9} (Figure 13). This novel binding specificity might be the result of an exchange at position II of CBL2. Here, the chimeric variant carries an aspartate instead of a glutamate (E227D) which could allow the binding of the sterically demanding sulfoglycans. A similar case has already been shown in a previous study where a Epa1A^{E227A} variant exhibited an increased binding of sulfated glycans (Ielasi et al. 2014). Apparently, the exchange of large amino acid residues in CBL2 with smaller ones seems to lead to an enhanced binding of sulfoglycans, possibly due to less obstruction of the sulfate group. Generally, the structural analysis of Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} supports the initial hypothesis that additional structural elements of the EpaA domain are involved in programming ligand binding specificity since an exchange of whole CBL2 motifs does not completely transfer binding specificity. Nevertheless, subtle changes in the composition or orientation of a CBL2 motif can change ligand binding specificity which underlines the importance of this structural motif for Epa functionality.

3.2 The variable L1 region influences the affinity of ligand binding

A central aspect of this study was the analysis of the variable loop L1 which is located in the outer binding pocket of Epa adhesion domains. Therefore, a mutational analysis of several EpaA domains was performed using a directed approach, because to date the function of this loop has not been addressed experimentally. Interestingly, most members of the Epa family carry a short L1 region that consists of four to seven amino acids while in four cases, Epa9, Epa10, Epa20 and Epa23, the loop is considerably longer and spans between 22 and 30 amino acids (Figure 8). A previous study showed that it is not possible to resolve the structure of the elongated L1 region by X-ray crystallography because the elongated loop appeared to be too flexible and was not defined by electron density (Diderrich 2014). In this work it could be demonstrated that the variable length of the L1 region has a functional role in host cell binding. This was achieved by the analysis of six chimeric EpaA domains that were created by exchange of the L1 regions in Epa1A, Epa6A, Epa9A and Epa10A. The identification of two gain-of-function mutants, Epa9A^{L1Epa1} and Epa10A^{L1Epa1}, and the loss-of-function variant Epa6A^{L1Epa9} clearly showed that a shortened L1 region can increase host cell adhesion while an extended L1 region corresponds with a weak host cell binding (Figure 10). In general this change in binding behavior could be caused by either modification of ligand binding affinity or by alteration of binding specificity. While the former would suggest a regulation of ligand access to the binding pocket, the latter likely involves the presentation of additional binding sites.

One possibility for the regulation of ligand access to the binding pocket could be a lid functionality of the elongated L1 region (Figure 32). Examples for lid-regulated substrate binding have been found in lipases (Xiao and Lowe 2015) and other enzymes that for example show transferase or hydrolase activity (Calvio et al. 2018; Del Caño-Ochoa et al. 2018), but to date this functionality has not been observed for lectins. If a long L1 region generally restricts ligand access to the inner binding pocket by acting as a lid, a reduced *in vitro* binding would be expected in addition to the reduced *in vivo* functionality observed in host cell binding assays. However, in this work a strong ligand binding with K_D values in the lower micromolar range was found for all EpaA variants regardless of L1 length (Figure 15). This finding is supported by the results of the glycan array analysis that showed strong binding for all L1 exchange variants. Finally, MD simulations with models of the Epa9A domain showed no occlusion of the inner binding pocket but either random movement of the L1 region or a fixation of loop L1 by attachment to loop L2. Taking these observations into consideration it can be concluded that the results of *in vitro* and *in silico* experiments do not support the hypothesis of a lid functionality. Alternatively, the conditions used for these tests could favor an ‘open’ conformation while *in vivo* adhesion measurements on human epithelial cells show a ‘closed’ state. This seems plausible because lid movement is energetically expensive (Stank et al. 2016) and thus could be hampered by the conditions of the MD simulations and FTS measurements. Another possibility would include fine tuning of the binding pocket by length or composition of the L1 region that could allow to program the binding of different ligand species. Here, the additional binding of sulfogalactosides

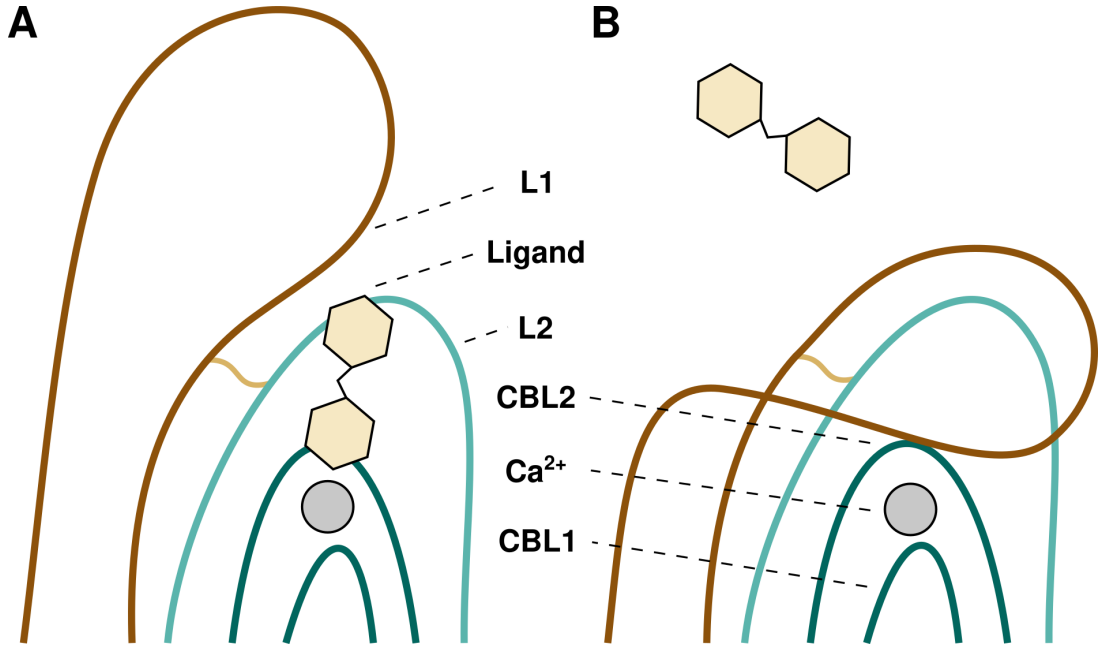

Figure 32: Scheme of putative L1 lid functionality.
A. Schematic view of an EpaA binding pocket carrying an elongated L1 region (brown) which is connected with the loop L2 (cyan) via a disulfide bond (light brown). A calcium ion (grey) is complex in the inner binding pocket. The conformation of loop L1 corresponds to an 'open' state that allows ligand (beige) access to the inner binding pocket (dark cyan). *B*. Host cell adhesion assays suggest that the long L1 region interferes with *in vivo* functionality of Epa9A and Epa10A. This observation can be rationalized by a lid functionality of loop L1 which changes conformation to occlude the inner binding pocket and thus prevent ligand binding.

by Epa9A^{L1Epa1} could be induced by such long range effects. However, these effects are hard to rationalize because they are difficult to detect by structural analysis (Veelders et al. 2010). For further analysis of these effects a variation of L1 length could be beneficial, as it would possibly show either a gradual change of ligand binding that is caused by long range effects or a switch between strong and weak binding caused by a lid functionality.

To address the question whether an exchange of L1 regions alters the ligand binding specificity of EpaA domains, chimeric L1 exchange variants were characterized by glycan array experiments. Here, a detailed analysis of ligand binding patterns revealed that the L1 region only has a minor impact on ligand binding specificity and preserved the binding patterns of the respective recipient variants. Nevertheless, the binding patterns found for Epa1A^{CBL2Epa9} and Epa9A^{CBL2Epa1} reveal novel best bound ligands showing that the L1 region is either directly or indirectly involved in the programming of ligand binding specificity (Figure 33). Only one of the five mutants, Epa9A^{L1Epa1}, showed a binding pattern with additional specificity toward sulfated galactosides and thus differed from its recipient. The appearance of a novel sulfoglycan specificity has already been demonstrated for Epa adhesion domains with mutated CBL2 motifs showing that a number of different mutations can result in the binding of sulfogalactosides (Diderrich et al. 2015; Ielasi et al. 2014). In the case of Epa9A^{L1Epa1} the additional sulfoglycan specificity may explain an increased host cell binding because the surface of cancer cells presents an increased amount of sulfated carbohydrates (Tanaka-Okamoto et al. 2017). The analysis of ligand binding patterns

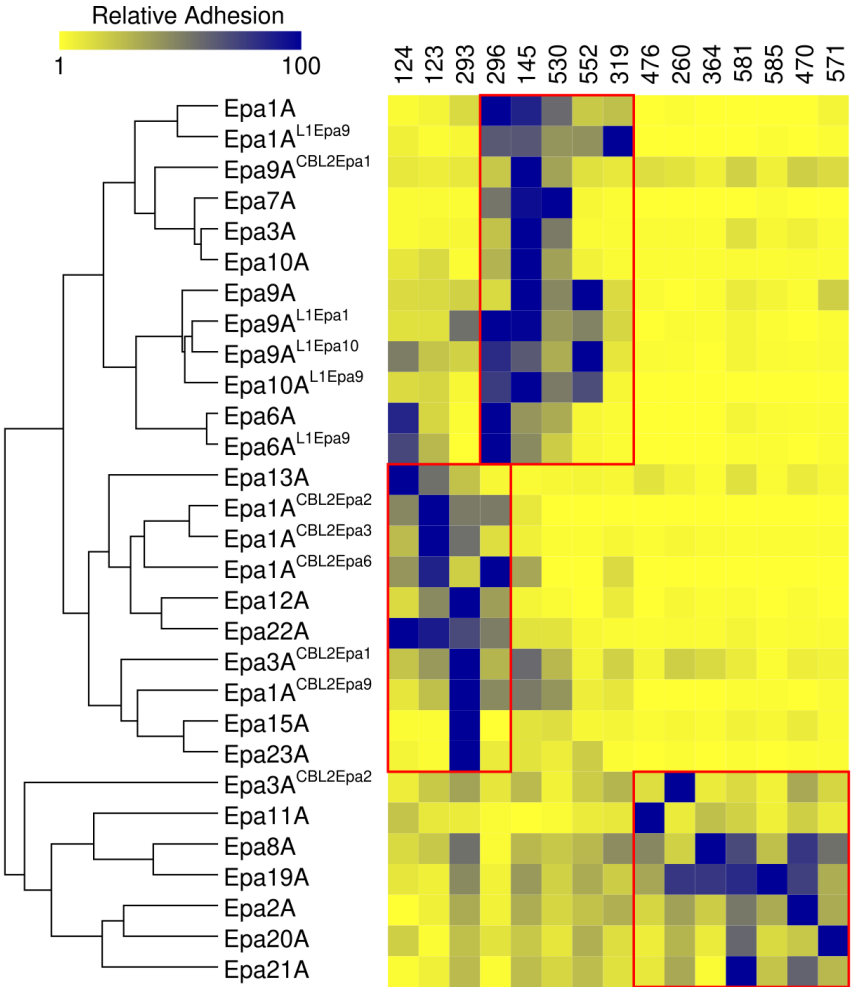

Figure 33: Cluster analysis of glycan array data showing the best bound ligands for each EpaA variant. Reprogramming of ligand binding specificity is visualized by the change of best bound ligands in glycan array analysis. Binding patters for the best bound glycan structures of each EpaA variant were taken from Figure 12. The comparison shows that for the L1 exchange variants the best bound ligands remain close to their respective recipient variants while for the CBL2 exchange variants a donor or novel behavior can be found (red boxes). The CFG number of each ligand is shown above.

also showed that in comparison with Epa1A, Epa3A and Epa6A the two variants Epa9A and Epa10A show a 1.8 times higher binding frequency toward long and complex glycan structures of at least eight moieties. Thus, the length of the L1 region seems to correlate with the length and complexity of their target carbohydrate structures. This finding is corroborated by MD simulations with Epa9A using gangliotetraose (Galβ1-3GalNAcβ1-4Galβ1-4Glc) as a ligand. Here, a direct interaction between the elongated L1 region and the glucose moiety of the ligand could be found (Figure 21). For further investigation of this finding, crystals of Epa9A were soaked with glycan 319 of the CFG glycan array (Figure 13B), a branched ligand that has kindly been provided by Prof. Dr. Carlo Unverzagt. Previous soaking attempts with disaccharide ligands successfully yielded co-crystal structures of Epa9A in complex with lacto-*N*-biose (Galβ1-3GlcNAc) and N-acetyl-D-lactosamine (Galβ1-4GlcNAc) which showed that Epa9A crystals generally are stable enough for soaking (Diderrich 2014). In this case, however, the crystals partly dissolved so no diffraction could be detected. To address this issue, gangliotetraose could be used for soaking or

co-crystallization since the simulation results suggest an efficient binding with direct contact between loop L1 and the ligand.

The present study could show that the L1 region mainly has an effect on the *in vivo* binding strength with a weak host cell adhesion generally being correlated with a long L1 loop. As there are other EpaA variants that also carry elongated L1 regions and exhibiting a weak adhesion strength, namely Epa20A and Epa23A, it would be interesting to also construct L1 exchange variants with these domains. This could provide useful information about a possible lid functionality or ligand binding function of the L1 and also shed light on the role of the L1 region in the determination of ligand binding specificity because these variants completely differ from Epa9A and Epa10A in their ligand binding patterns (Diderrich et al. 2015). Here it could be analyzed whether these exchange variants also exhibit an increased target cell binding. In another approach it could be investigated how the length of an L1 region translates into a specific host cell adhesion by creating a number of different mutants carrying shortened loops. This could be especially interesting because Epa20A and Epa23A carry shorter L1 regions than Epa9A or Epa10A but also exhibit weak target cell binding. In addition to a more detailed *in vivo* characterization of L1 mutants, it could also be attempted to structurally analyze different conformations of loop L1. Unfortunately, a characterization of the 'open' conformation by X-ray crystallization seems not feasible since for Epa1A it has already been shown that crystallization was not possible without a ligand (Maestre-Reyna et al. 2012). To address this issue it might be possible to use hydrogen–deuterium exchange mass spectrometry (HDX-MS) because the 'open' and 'closed' states should result in different deuteration of the protein and thus could show the proposed lid functionality.

3.3 CBL2 and L1 have distinct but different effects on EpaA functionality

In this study it was shown that replacing only the CBL2 motif or the L1 region is not sufficient to successfully transfer the binding behavior of an EpaA donor domain to a recipient domain. Instead, the *in vivo* and *in vitro* characterization of different exchange variants suggests that more than one variable structural element is included in programming the binding functionality. This hypothesis was tested by the analysis of three double exchange variants that carried both elements, CBL2 and L1, of a particular donor variant. Two of the three variants, Epa9A$^{CBL2+L1Epa1}$ and Epa10A$^{CBL2+L1Epa1}$, show a host cell adhesion that is lacking the strong host cell binding of their respective L1 mutants. Instead, both double exchange mutants exhibit weak binding behavior comparable to their corresponding CBL2 exchange variants (Table A1). Furthermore, the strong host cell binding of Epa1A was only reduced in the double exchange variant Epa1A$^{CBL2+L1Epa9}$ but not in Epa1A^{CBL2Epa9} or Epa1A^{L1Epa9}. Overall, this shows that several structural elements are effectively involved in programming the binding behavior but also reveals that neither the host cell binding nor the ligand binding specificity of the Epa family is fully controlled by CBL2 and L1. The findings instead suggest that the CBL2 motif is of central importance in defining ligand binding specificity while loop L1 appears to primarily affect binding affinity. At least

one additional factor seems to be also involved in programming functionality because only one double exchange variant showed an actual donor behavior. However, discovering additional factors that influence the binding behavior might be an intricate task as for example long range effects can be hard to detect by structural analysis (Veelders et al. 2010).

In contrast to the weak *in vivo* binding to human epithelial cells, *in vitro* FTS measurements with the three double mutants revealed strong T-antigen binding with K_D values in the lower micromolar range (Figure 15). This interesting finding is supported by the results of several glycan array analyses that have revealed specific binding of T-antigen for Epa1A, Epa9A and Epa10A as well as their single exchange variants (Maestre-Reyna et al. 2012; Diderrich et al. 2015). As this strong binding to T-antigen is still present in the double exchange variants, it is obvious that even a combined exchange of both variable motifs does not abrogate the overall functionality of an EpaA domain. Nevertheless, the binding of other ligands might be impaired in these novel chimeric variants which could explain the reduced *in vivo* binding. To investigate this possibility, a glycan array analysis of the novel double mutants would be necessary which would give further insights about the dominant role of the CBL2 motif. Alternatively, FTS measurements with additional α- and β-linked galactosides could show whether the double mutants still discriminate between glycan structures with different linkage types and thus exhibit recipient or donor behavior.

3.4 Sulfoglycan binding is mediated by conserved structural elements

Another part of this study is dedicated to the sulfoglycan-binding specificity of Epa12, Epa15, Epa22 and Epa23, which have not yet been analyzed in detail. In the case of Epa22A, host cell binding has not been tested before, so in this work it was measured for the first time revealing that Epa22A mediates a very low adhesion strength (Figure 10). This could also be confirmed by *in vitro* measurements with mono- and disaccharides showing an up to 13 times lower binding strength than other natural EpaA variants (Figure 25). Here, the overall weak binding of Epa22A indicates a different function than adhesion to human epithelial cells. It has already been shown that the production of Epa22 is upregulated in biofilms, upon osmotic stress and glucose starvation (Kraneveld et al. 2011; Roetzer et al. 2008). These conditions can be found in later stages of an infection which suggests other targets like neighboring cells or cells of the host immune system since fungal pathogens are known to evade or even survive the ingestion by macrophages (Kasper et al. 2015). Furthermore, the preferential binding of sulogalactosides could hint toward a connection between Epa22 and immune cells because previous studies have shown that sulfomucins are linked to lymphocytes, an important part of the immune system (Nieuw Amerongen et al. 1998). Evidence for specific functions of EpaA domains besides host cell adhesion has also been found in other cases, including adhesion to hydrophobic surfaces or fibronectin (Epa1, Epa6 and Epa7), macrophages (Epa1) and *C. albicans* hyphae (Epa8 and Epa19) during oral infection (Kuhn and Vyas 2012; Valotteau et al. 2019; Tati et al. 2016). As infections with *C. glabrata* often involve other fungal or bacterial species (Klotz et al. 2007) it

seems probable that Epa22A could have a function in the interaction with other pathogens during mixed infections. Additionally, other PA14-like adhesins like Flo1 and the flocculin Flo5 of *S. cerevisiae* are known to mediate self-adhesion (Brückner and Mösch 2012), a function that is important for the formation of biofilms. In *C. glabrata* however, it remains to be determined which adhesins are involved in self-adhesion and biofilm formation. Here, glycan array experiments using microbe-focused glycan libraries could help to investigate binding of Epa22A to other pathogens (Geissner et al. 2019).

To date, it is not fully understood which combination of structural features inside an EpaA domain is responsible for programming a specific sulfoglycan binding, because the four natural variants Epa12A, Epa15A, Epa22A and Epa23A differ in both, CBL2 composition and L1 length (Figure 8). Based on a number of previous glycan array analyses it has been suggested that the sulfoglycan specificity of Epa12A and Epa15A could be controlled by a tyrosine at position III of CBL2 that possibly interacts with 6-sulfated terminal galactose moieties via its hydroxy group (Diderrich 2014). Similarly, the lysine at CBL2 position IV might be responsible for a comparable interaction in Epa23A. Here, these putative interactions were investigated by mutation of the CBL2 motif and subsequent analysis of *in vivo* adhesion to human epithelial cells. It appeared that in general the *in vivo* adhesion of Epa12A^{Y227F}, Epa15A^{Y227F} and Epa23A^{K231A} is reduced compared to their respective natural variants (Figure 23) suggesting that positions III and IV of CBL2 could be involved in sulfoglycan binding by interaction of their polar side chains with the sulfogalactose moiety (Maestre-Reyna et al. 2012). However, it has already been shown in this work that in most cases a mutation of the CBL2 motif leads to a reduced *in vivo* adhesion (Figure 10). Therefore, the question whether CBL2 positions III and IV are involved in sulfoglycan binding cannot be conclusively answered by just analyzing host cell binding. To address this issue, all three novel mutants were tested for *in vitro* binding to different sugars using fluorescence titration spectroscopy. In general, the titration results were consistent with the *in vivo* adhesion assays by showing a slightly lower binding strength of the mutants compared to their respective natural variants, which supports the hypothesis that the mutated residues are important for the binding of sulfated glycans. A surprising exception to this finding was the increased binding strength of Epa15A^{Y227F} toward 6-sulfogalactose which contradicts the initial hypothesis. In fact, either a reduced or unchanged adhesion would have been expected in this case. Based on the available glycan array data it was expected that the tested natural EpaA domains would show a clear preference for 6-sulfogalactose. The lack of such a preference indicates that the tested Epa variants may not be able to clearly distinguish between sulfated and nonsulfated monosaccharides, but rather require the presence of at least one additional hexose unit. This conclusion is in agreement with previously obtained structural data of Epa1A in complex with 6-sulfogalactose that did not show direct interactions between the sulfate group of the ligand and Epa1A (Kock 2015). To exclude the possibility that the missing discrimination is only limited to monosaccharide ligands, longer sugars consisting of at least two to three moieties would be needed for further analysis.

An additional aim of this study was to identify the structural basis for sulfoglycan specificity by the analysis of protein crystals. Therefore, a number of crystallization experiments with Epa12A, Epa15A and Epa23A were performed using either lactose or 6-sulfogalactose as a ligand. In the case of Epa12A and Epa23A crystals were successfully produced, but these did not provide usable diffraction data. Obviously, the crystallization conditions in both cases were not appropriate for the growth of high quality crystals and need further optimization. As crystalline growth could already be observed, the next step would be to perform seeding experiments for the production of crystals that are suitable for structural characterization. The lack of crystals for Epa15A likely can be attributed to the low purity of the protein fractions (Figure 24). Here, it could be beneficial to try the expression of different Epa15A fusion proteins to improve the final purity of the eluted protein. Possible fusion proteins could include NusA or other tags like FLAG or STREP II because these are reported to yield a considerably higher purity than the HIS tag used in the present study (De Marco et al. 2004; Lichty et al. 2005).

Despite the lack of suitable crystal structures, an attempt was made to find possible binding modes for sulfoglycans by modeling all four EpaA variants and performing docking simulations with [6S]Galβ1-3[6S]GlcNAc, a sulfoglycan structure that showed the strongest binding in initial glycan studies with natural adhesion domains (Figure 13). Here it was found that for Epa12A, Epa15A and Epa23A the sulfate groups of the ligand showed interactions with conserved residues like the aspartates of the D*cis*D motif, the arginine at position I of CBL2 or the tryptophan residue in L3. Similar interactions were found in docking simulations with Epa1A^{CBL2Epa9} and [6S]Galβ1-3[6S]GlcNAc. This raises the question why other Epa variants do not show specific binding of sulfated galactosides because these residues are conserved in most members of the Epa family (Figure 8). A possible answer to this question might be provided by glycan array experiments with CBL2 exchange variants shown in Figure 13 and by the results of a previous study analyzing the structure of Epa1A in complex with 6-sulfogalactose (Kock 2015). The former shows that an exchange of CBL2 motifs can lead to an increased binding of sulfated galactosides while the latter suggested that EpaA binding domains might generally have the capability of binding derivatized glycan structures and that this specificity is narrowed by modification of variable structural elements like the CBL2 motif. If this is the case, it would represent an important mechanism for the evolution of pathogenicity in *C. glabrata*. To test this hypothesis, systematic mutation studies would be necessary that target the CBL2 motif of different EpaA domains. As to date only the CBL2 motif Epa1A has been addressed by random mutagenesis experiments, it could be beneficial, to also use this approach for other EpaA domains to reveal, if generally many CBL2 sequences favor the binding of sulfated galactosides. In general, the initial hypothesis about an interaction between CBL2 positions III and IV and sulfated ligand structures could not be confirmed but the results of the present study suggest that a specificity for sulfated glycan structures is mainly given by conserved residues. Furthermore, it seems plausible that a promiscuous binding of hexoses carrying functional groups at C3 or C6 is beneficial for the development of more specialized glycan binding by the modification of variable regions of the binding pocket.

4 Outlook

The present study shows that the programming of Epa functionality is a complex exercise. Basic Epa functionality is given by conserved structural elements that (i) maintain the conformation of the binding pocket, (ii) coordinate a calcium ion for direct interaction with carbohydrate ligands and (iii) provide galactose binding specificity by stacking interactions (Maestre-Reyna et al. 2012). In this work it was also found that the binding of sulfated galactosides could also be mainly dependent on conserved residues of the EpaA binding pocket. Furthermore, ligand binding specificity and host cell binding are mainly determined by a combination of variable structural elements inside the binding pocket. By uncovering the structural basis for sulfogalactose binding, it could be possible to manufacture molecular probes which specifically target this type of glycans and thus could be used as molecular probes for the identification of cancer cells, which are known to present an unusually high amount of sulfoglycans (Tanaka-Okamoto et al. 2017). In addition to the basic functionality given by conserved residues, the CBL2 motif largely controls the programming of ligand binding specificity while the L1 region affects binding affinity by either direct contact to the ligand or by indirect structural effects (Diderrich et al. 2015). Here, the length of loop L1 has been shown to be especially important because it directly influences the *in vivo* binding behavior of an Epa adhesion domain. In this work it was also found that the elongated loop L1 could show a lid functionality. If such a functionality could be confirmed, it would provide the first example of a lid-controlled binding pocket in a lectin.

In addition to the two variable elements CBL2 and L1 it is likely that further factors are involved in the determination of binding specificity because a given ligand binding pattern cannot be completely transferred by the exchange of these structural hot spots. However, as these factors can include a number of different parameters like loop dynamics and hydration effects, their influence on Epa functionality might be difficult to determine. To further analyze the large family of Epa proteins it could be beneficial to employ additional procedure like random mutagenesis which has been shown to be an effective approach (Ielasi et al. 2014). Therefore, a directed evolution of Epa-like adhesins using a combination of random mutagenesis and careful selection of mutants with specific binding patterns could allow to gradually program a given ligand binding profile. The ability to produce lectins with specific binding characteristics would be a great advantage in the challenging task of deciphering the sugar code (Hu et al. 2012; Hu et al. 2015). These lectins could be useful for the analysis of biological surfaces and also could serve as useful tools for medical diagnostics or glycan profiling. Specifically, they could be used for the analysis of complex carbohydrate structures on mammalian cells or microbial surfaces (Smith and Cummings 2013).

In the case of Epa22A it would be interesting to further investigate its adhesion functionality as it does not mediate adhesion to human cells and has a weaker binding of T-antigen than other EpaA variants. Here, it could be tested, whether Epa22A mediates adhesion between *C. glabrata* and other pathogens in mixed infections like it has already been shown for Epa8A and Epa19A

(Tati et al. 2016). This could provide novel insights into the dynamics of mixed infections and alleviate the development of suitable treatments. Another possible function of Epa22A could be a role in the colonization of habitats different from the human body. As *C. glabrata* frequently can be isolated from other places like coffee beans or the cloaca of seagulls (de Melo Pereira et al. 2014; Al-Yasiri et al. 2016) it seems plausible that several EpaA domains that do not show adhesion to human cells are involved in the colonization of other habitats. For this reason it could be beneficial to analyze the abundance of different Epas in strains that were not isolated from the human body but from other places.

5 Materials

5.1 Chemicals, enzymes, kits and other material

Unless stated otherwise, the chemicals used in this work have been purchased from the companies *Carl Roth GmbH + Co. KG* and *Sigma Aldrich*. All plastic material and filters were purchased from either *Merck Millipore* or *Sarstedt AG & Co. KG*. Complex glycans for soaking of protein crystals were kindly provided by the lab of Prof. Dr. Carlo Unverzagt (University of Bayreuth, Germany). Additional chemicals and materials are listed in Table 7.

Table 7: Chemicals and materials

Description	Source
96-well Innovaplate	*Agilent*
Alexa Fluor 488 TFP ester	*Thermo Fisher Scientific*
Alexa Fluor 647 TFP ester	*Thermo Fisher Scientific*
CryoLoops	*Hampton Research*
Crystal Wand	*Hampton Research*
D-galactose-6-sulfate	*Dextra Laboratories*
Dulbecco's Modified Eagle Medium	*Thermo Fisher Scientific*
E.Z.N.A. Plasmid DNA Mini Kit	*Omega Bio-tek*
E.Z.N.A. Gel Extraction Kit	*Omega Bio-tek*
GeneRuler DNA Ladder Mix	*Thermo Fisher Scientific*
GelRed	*Biotium*
HiLoad Superdex 75 prep grade column	*GE Healthcare*
MicroMounts	*MiTeGen*
NEXTAL Screening Suites	*Qiagen*
NucleoSpin Plasmid kit	*Macherey-Nagel GmbH & Co. KG*
Omnifix Luer Solo Syringe 1 ml, 5 ml and 20 ml	*B. Braun*
PageRuler Prestained Protein Ladder	*Thermo Fisher Scientific*
PD-10 Desalting Columns	*GE Healthcare*
Phusion High-Fidelity DNA Polymerase	*Thermo Fisher Scientific*
Protino Ni-NTA Columns	*Macherey-Nagel GmbH & Co. KG*
T-Antigen (Galβ1-3GalNAc)	*Santa Cruz Biotechnology*
T4 DNA Ligase	*Roche Applied Science*
Trypsin	*Thermo Fisher Scientific*
Unipuck	*MiTeGen*
Venor GeM One Step	*Minerva Biolabs*
VIEWseal	*Greiner Bio-One*

5.1.1 Media, buffers and antibiotics

Media were prepared according to Table 8. For solid media the solution was supplemented with 2 % agar. If necessary, glucose was added after preparation and autoclaving. Antibiotics were ordered from *Carl Roth GmbH + Co. KG* and used at the concentrations given in Table 9. All media were autoclaved after preparation and prior to the addition of antibiotics. Buffers used for protein preparation were sterile filtered and degassed.

Table 8: Composition of media and buffers

Solution	Amount	Component
AM(+) buffer	20 mM	Tris pH 8.0
	200 mM	NaCl
	(5 mM)	($CaCl_2$)
AML buffer	20 mM	Tris pH 8.0
	200 mM	NaCl
	50 mM	Lactose
	5 mM	$CaCl_2$
Amino acid dropout mix	2 g	of every standard L-amino acid
	2 g	Adenine
	0.2 g	4-aminobenzoic acid
Coomassie Brilliant Blue destaining solution	5 %	Acetic acid
	50 %	Ethanol
	45 %	H_2O
Coomassie Brilliant Blue staining solution	0.1 %	Coomassie Brilliant Blue G250
	0.1 %	Coomassie Brilliant Blue R250
	5 %	Acetic acid
	50 %	Ethanol
	45 %	H_2O
DNA loading dye (6 ×)	10 mM	Tris pH 7.6
	0.03 %	Bromophenol blue
	0.03 %	Xylene cyanol
	60 mM	EDTA, pH 8.0
	60 %	Glycerol

Continuation on next page.

Solution	Amount	Component
High urea buffer (HU buffer)	8 M	Urea
	200 mM	NaH_2PO_4 pH 6.8
	5 % (w/v)	SDS
	0.1 mM	EDTA pH 8.0
	0.03 % (w/v)	Bromophenol blue
	100 mM	DTT (freshly added)
Labeling buffer	50 mM	$NaHCO_3$
LB medium ("lysogeny broth")	1 % (w/v)	Trypton
	0.5 % (w/v)	Yeast extract
	1 % (w/v)	NaCl
LFM ("Low Fluorescence Medium")	10 % (v/v)	Salt stock
	0.1 % (v/v)	Vitamin stock
	0.1 % (v/v)	Trace element stock
	2 % (w/v)	Glucose
PBS buffer	137 mM	NaCl
	2.7 mM	KCl
	10 mM	Na_2HPO_4
	1.8 mM	KH_2PO_4
PEG buffer	100 mM	Lithium acetate
	10 mM	Tris pH 8.0
	1 mM	EDTA pH 8.0
	40 % (w/v)	PEG 4000
RF1	100 mM	RbCl
	50 mM	$MnCl_2$
	30 mM	KOAc pH 7.5
	10 mM	$CaCl_2$
	15 %	Glycerol
RF2	10 mM	RbCl
	10 mM	MOPS pH 6.8
	75 mM	$CaCl_2$
	15 %	Glycerol

Continuation on next page.

Solution	Amount	Component
SC-Ura medium (“Synthetic Complete” without L-Uracil)	0.15 % (w/v)	“Yeast Nitrogen Base”
	0.5 % (w/v)	Ammonium sulfate
	0.2 mM	myo-Inositol
	0.2 % (w/v)	Amino acid dropout mix (-Ura)
Salt stock	1 % (w/v)	KH_2PO_4
	0.5 % (w/v)	$MgSO_4$
	0.1 % (w/v)	NaCl
	0.1 % (w/v)	$CaCl_2$
	5 % (w/v)	$(NH_4)_2SO_4$
SDS runnning buffer (10 ×)	30 g	Tris
	144 g	Glycine
	10 g	SDS
SORB buffer	100 mM	Lithium acetate
	10 mM	Tris pH 8.0
	1 mM	EDTA pH 8.0
	1 M	Sorbitol
Tace element stock	0.05 % (w/v)	H_3BO_4
	0.004 % (w/v)	$CuSO_4$
	0.01 % (w/v)	KI
	0.02 % (w/v)	$FeCl_3$
	0.04 % (w/v)	$MnSO_4$
	0.02 % (w/v)	Na_2MoO_4
	0.04 % (w/v)	$ZnSO_4$
TAE buffer	40 mM	Tris
	20 mM	NaOAc
	2 mM	EDTA
Vitamin stock	0.0002 % (w/v)	Biotin
	0.04 % (w/v)	Calciumpantothenate
	0.2 % (w/v)	myo-Inositol
	0.04 % (w/v)	Niacin
	0.02 % (w/v)	Para-aminobenzoic acid
	0.04 % (w/v)	Pyroxidin-HCl
	0.04 % (w/v)	Thiamine-HCl

Continuation on next page.

Solution	Amount	Component
YEPD medium	2 % (w/v)	Trypton
	1 % (w/v)	Yeast extract
	2 % (w/v)	Glucose

Table 9: Antibiotics

Solution	Stock concentration	Final concentration	Supplier
Ampicillin	100 mg/ml	100 µg/ml	*Carl Roth GmbH + Co. KG*
Kanamycin	35 mg/ml	35 µg/ml	*Carl Roth GmbH + Co. KG*
Penicillin	10 000 U/ml	100 U/ml	*Carl Roth GmbH + Co. KG*
Streptomycin	10 mg/ml	100 µg/ml	*Carl Roth GmbH + Co. KG*

5.1.2 Antibodies

The antibodies in Table 10 were used for the preparation of cells for fluorescence microscopy and *in vivo* adhesion assays on human epithelial cells. Antibodies were used at a dilution of 1:1000 in PBS buffer with 1 % BSA. Labeled antibodies were conjugated with either Cy3 or Dylight488.

Table 10: Antibodies

Name	Description	Source
Mouse anti-HA	monoclonal	*Sigma Aldrich*
Goat anti-mouse	polyclonal, Cy3 labeled	*Sigma Aldrich*
Rabbit anti-*Candida*	polyclonal	*OriGene Technologies GmbH*
Goat anti-rabbit	polyclonal, Dylight488 labeled	*Dianova*

5.1.3 Devices and Machines

Table 11: Devices

Device	Source
Autoclave “LTA 2x3x4”	*Zirbus Technology GmbH*
CCD camera “1394 ORCA-ERA”	*Hamamatsu Photonics Deutschland GmbH*
Centrifuge “4K-14”	*Sigma Laborzentrifugen GmbH*

Continuation on next page.

Device	Source
Centrifuge “Beckmann L7-65 Ultracentrifuge”	*Beckman Coulter Life Sciences*
Centrifuge “Biofuge pico”	*Heraeus Holding GmbH*
Centrifuge “Megafuge 1.0R”	*Heraeus Holding GmbH*
Centrifuge “Sorvall RC-5B Plus Superspeed”	*Thermo Fisher Scientific*
Chromatography system “ÄKTA Purifier UPC 10”	*GE Healthcare*
Cold light source “KL 1500 LCD”	*Schott AG*
Cold trap “CT 02-50”	*Martin Christ Gefriertrocknungsanlagen GmbH*
Digital camera “Canon Power Shot A650”	*Canon Deutschland GmbH*
DNA gel documentation system “Gel Doc XR”	*Bio-Rad Laboratories GmbH*
DNA gel electrophoresis chambers	workshop, Philipps-Universität Marburg
Fluorescence microscope “Axiovert 200M”	*Carl Zeiss AG*
French pressure cell press	*American Instrument Company*
Fluorimeter “Infinite 200 Pro”	*Tecan*
Incubation shaker “Multitron Pro”	*Infors*
Incubator “B5042”	*Heraeus Holding GmbH*
Incubator “Heracell 150i”	*Thermo Fisher Scientific*
Homogenizer “MP Bio FastPrep 24”	*Kem-En-Tec Nordic*
Peristaltic pump “Pump P-1”	*Pharmacia Biotech*
pH meter “S20 Seven Easy pH”	*Mettler-Toledo GmbH*
Pipetting robot “Cartesian Microsys SQ 4004”	*Genomic Solutions*
Power supply for DNA gel electrophoresis “PowerPac Basic”	*Bio-Rad Laboratories GmbH*
Protein gel electrophoresis system “Mini Protean 3 Cell”	*Bio-Rad Laboratories GmbH*
Protein crystallization imager “Rock Imager”	*Formulatrix*
Real-time thermocycler “CFX Connect Real-Time PCR Detection System”	*Bio-Rad Laboratories GmbH*
Roller “Stuart SRT1”	*Bibby Scientific Limited*
Rotation incubator	workshop, Philipps-Universität Marburg
Sonicator “Sonopuls HD 3100”	*Bandelin Electronic GmbH & Co. KG*
Spectrofluorometer “FP-6500”	*Jasco*

Continuation on next page.

Device	Source
Spectrophotometer “NanoDrop ND-1000”	*Thermo Fisher Scientific*
Spectrophotometer “Ultrospec 10”	*Amersham Biosciences*
Stereo microscope “Stemi 2000-C”	*Carl Zeiss AG*
Thermocycler “Primus 25 advanced”	*VWR International*
UV transilluminator “IL-200-M/L”	*H. Saur Laborbedarf*
Vaccuum pump “MZ 2C”	*Vacuubrand GmbH & Co. Kg*
Vortex mixer “Vortex-Genie 2”	*Scientific Industries*

5.1.4 Software

Table 12: Software

Software	Source
Clone Manager 9	*Sci-Ed Software*
Clustal Omega	https://www.ebi.ac.uk/Tools/msa/clustalo/ (Madeira et al. 2019)
ConSurf	https://consurf.tau.ac.il/ (Glaser et al. 2003; Landau et al. 2005)
ExPASy Prot Param	Swiss Institute of Bioinformatics (Gasteiger et al. 2005)
Fiji	https://www.fiji.sc (Arganda-Carreras et al. 2017)
Jalview	https://www.jalview.org
Morpheus	https://software.broadinstitute.org/morpheus
Quantitiy One 1-D	*Bio-Rad*
PyMOL	*Schrödinger*
QtiPlot	*Iondev SRL*
SnapGene Viewer	*GSL Biotech*
UCSF Chimera	http://www.rbvi.ucsf.edu/chimera
Unicorn 5.0	*Amersham plc*
Volocity 3D	*Perkin Elmer*

5.1.5 Sources of supply

- *Agilent* (Santa Clara, USA)
- *American Instrument Company* (Silver Spring, USA)
- *Amersham Biosciences* (Amersham, United Kingdom)
- *Amersham plc* (Amersham, United Kingdom)
- *B. Braun* (Melsungen, Germany)
- *Bandelin Electronic GmbH & Co. KG* (Berlin, Germany)
- *Beckman Coulter Life Sciences* (Brea, USA)
- *Bibby Scientific Limited* (Stone, United King-

dom)
- *Bio-Rad Laboratories GmbH* (Hercules, USA)
- *Biotium, Inc.* (Fremont, USA)
- *Canon Inc.* (Tokyo, Japan)
- *Carl Roth GmbH + Co. KG* (Karlsruhe, Germany)
- *Carl Zeiss AG* (Oberkochen, Germany)
- *Dextra Laboratories* (United Kingdom)
- *Dianova* (Hamburg, Germany)
- *Formulatrix* (Bedford, USA)
- *GE Healthcare* (Chicago, USA)
- *Greiner Bio-One* (Kremsmünster, Austria)
- *GSL Biotech* (Chicago, USA)
- *H. Saur Laborbedarf* (Reutlingen, Germany)
- *Hamamatsu Photonics K.K.* (Hamamatsu City, Japan)
- *Hampton research* (Alison Viejo, USA)
- *Heraeus Holding GmbH* (Hanau, Germany)
- *Infors AG* (Basel, Switzerland)
- *Iondev SRL* (Bukarest, Romania)
- *Jasco* (Easton, USA)
- *Kem-En-Tec Nordic* (Tåstrup, Denmark)
- *Macherey-Nagel GmbH & Co. KG* (Düren, Germany)
- *Martin Christ Gefriertrocknungsanlagen GmbH* (Osterode am Harz, Germany)
- *Merck Millipore* (Billerica, USA)
- *Mettler-Toledo GmbH* (Columbus, USA)
- *Microsynth AG* (Balgach, Switzerland)
- *Minerva Biolabs* (Berlin, Germany)
- *MiTeGen* (Ithaca, USA)
- *New England Biolabs* (Ipswitch, USA)
- *Omega Bio-tek, Inc.* (Norcross, USA)
- *OriGene Technologies GmbH* (Herford, Germany)
- *Perkin Elmer* (Waltham, USA)
- *Pharmacia Biotech* (Uppsala, Sweden)
- *Roche Applied Science* (Penzberg, Germany)
- *Santa Cruz Biotechnology, Inc.* (Dallas, USA)
- *Sarstedt AG & Co. KG* (Nümbrecht, Germany)
- *Schott AG* (Mainz, Germany)
- *Schrödinger* (New York, USA)
- *Sci-Ed Software* (Westminster, USA)
- *Scientific Industries* (Bohemia, USA)
- *Sigma Aldrich* (St. Louis, USA)
- *Sigma Laborzentrifugen GmbH* (Osterode am Harz, Germany)
- *Tecan (Männedorf, Switzerland)*
- *Thermo Fisher Scientific* (Waltham, USA)
- *Vacuubrand GmbH & Co. Kg* (Wertheim, Germany)
- *VWR International* (Radnor, USA)
- workshop, Philipps-Universität (Marburg, Germany)
- *Zirbus Technology GmbH* (Bad Grund, Germany)

5.2 DNA synthesis and sequencing

Synthesis of oligonucleotides for DNA amplification and sequencing of DNA was done by *Microsynth AG* (Balgach, Switzerland). Oligonucleotides used for PCR are shown in Table 13 with restriction sites in bold and mutated or inserted bases in lower case letters. Synthesis of the mutant *EPAA* domains $EPA1A^{L1EPA9}$, $EPA6A^{L1EPA9}$, $EPA10A^{L1EPA9}$, $EPA9A^{L1EPA1}$ and $EPA9A^{L1EPA10}$ for expression in *E. coli* was performed by *Thermo Fisher Scientific*. The respective fragments were then used for the creation of *E. coli* and *S. cerevisiae* expression plasmids.

Table 13: Primer

Name	Sequence (5'→3')
3152-SacI_rv	CTCAGAAGAATCGTAGCTGATTAAACC
3152-SacII_fw	**CCGCGG**ACATCTTCCAATGATATCAGTTTAGC
3153-SacI_rv	**GAGCTC**CGAAGTATCATAACTAACTAAGCCTG
3153-SacII_fw	**CCGCGG**AAGGATGACTATTCTTCCTCCTTG
3154-SacI_rv	**GAGCTC**ATCGTCACAGACAGTTCTGTATC
3154-SacII_fw	**CCGCGG**GACTTCACACCGTTTGC
3155/56-SacI_rv	**GAGCTC**ATCATCACAGACAGTCTTGTATCTG
3155/56-SacII_fw	**CCGCGG**GACATCACCCCGTTTGC
3xHA_seq_fw	CTATGACGTCCCGGACTATG
3xHA_seq_fw	CTATGACGTCCCGGACTATG
Epa10_CBL2-1_fw.AQ	CTACAACAACAGAGAATATGATTCTAAATTCAAGATC GGCTTCTAC
Epa10_CBL2-1_rv.AQ	ATCATATTCTCTGTTGTTGTAGAACAATCTCAATG
EPA12A_CBL2_YtoF_fw	CAATAACAGAGATttcTACGGTCAGATGAG
EPA12A_CBL2_YtoF_rv	CTCATCTGACCGTAgaaATCTCTGTTATTG
EPA12A_CBL2-1_fw	CAATAACAGAgaatacgacGGTCAGATGAGAC
EPA12A_CBL2-1_rv	GTCTCATCTGACCgtcgtattcTCTGTTATTG
EPA12A_NdeI_fw	**CATATG**CTCGAACCTGATAATTCCCTAC
EPA12A_XhoI_rv	**CTCGAG**ttaTGGGCATCCTTCAGAAGTATC
EPA15A_CBL2_YtoF_fw	CTACAATAATCGAGATtttATTGCTCAATTAGAC
EPA15A_CBL2_YtoF_rv	GTCTAATTGAGCAATaaaATCTCGATTATTGTAG
EPA15A_CBL2-1_fw	CTACAATAATCGAgaatatgatGCTCAATTAGAC
EPA15A_CBL2-1_rv	GTCTAATTGAGCatcatattcTCGATTATTGTAG
EPA15A_NdeI_fw	**CATATG**TATCATGATCCAACGACATTTCC
EPA15A_XhoI_rv	**CTCGAG**ttaTGGTGATTGTACATCTGCACATTG
EPA22A_CBL2_KtoA_fw	GTTTTTCAATAACATTGGTgcaGATTCTTCATTGGATC
EPA22A_CBL2_KtoA_rv	GATCCAATGAAGAATCtgcACCAATGTTATTGAAAAAC
EPA22A_CBL2-1_fw	GTTTTTCAATAACagagaatatGATTCTTCATTGGATC
EPA22A_CBL2-1_rv	GATCCAATGAAGAATCatattctctGTTATTGAAAAAC
EPA22A_SacI_rv	**GAGCTC**CTTGCATGTAGTATCATAGCCAA

Continuation on next page.

Name	Sequence (5'→3')
EPA22A_SacII_fw	**CCGCGG**AAACGAGCATTAAGTTCACAG
EPA23A_CBL2_KtoA_fw	CAATAATAGGGACTTTgcaGGTGCGCTG
EPA23A_CBL2_KtoA_rv	CAGCGCACCtgcAAAGTCCCTATTATTG
EPA23A_CBL2-1_fw	GTTTTACAATAATAGGgaatatgatGGTGCGCTGAATATG
EPA23A_CBL2-1_rv	CATATTCAGCGCACCatcatattcCCTATTATTGTAAAAC
EPA23A_L1-1_fw	cggaagaagggttcgtacccaTGCTGGGACCCATCATATC
EPA23A_L1-1_rv	tgggtacgaacccttcttccgATAACTATAGTTATATAGTTCCATA GTCAAGC
EPA23A_NdeI_fw	**CATATG**ATTCCGAACCCATTCGAGG
EPA23A_XhoI_rv	**CTCGAG**ttaTCCTGATACATTACCACATTCGG
L1-1_Epa10_fw.AQ	CGGAAGAAGGGTTCGTACCCATGTTGGAACCCAGAGT ACC
L1-1_Epa10_rv.AQ	TGGGTACGAACCCTTCTTCCGGTAAGGGTAGTTGTAA AGGTTCATG
L1-Epa10A_seq_fw	AATTGAAGGAAGGGCCGGTC
L1-EPA1A_seq_fw	CGGAAGAAGGGTTCGTACCC
L1-EPA9A_seq_fw	ACAGATAGTGATGCTGACGG
T7_seq_Pr_fw	TAATACGACTCACTATAGGG

5.3 Strains and plasmids

5.3.1 *E. coli* strains

Top10

Chemically competent *E. coli* Top10 (*Thermo Fisher Scientific*) were used for general cloning and plasmid propagation, because they have a high replication rate and a transformation efficiency. Additionally, the gene coding for endA endonuclease is deleted to prevent nonspecific plasmid digestion by endonuclease I and maximize the yield of plasmid preparations. Nonspecific recombination is reduced by a missense mutation in the *recA* gene.

Genotype: F^- *mcrA Δ(mrr-hsdRMS-mcrBC)* Φ80*lacZΔ*M15 *ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL* (Str^R) *endA1 nupG*

SHuffle T7 Express

The SHuffle T7 Express strain (*New England Biolabs*) is used for protein production under the control of the T7 expression system. It is especially useful for the production of proteins with

multiple disulfide bonds. Therefore it carries the DsbC chaperone, as well as deletions in the *trxB* and *gor* genes, which allows efficient formation of disulfide bonds in the cytoplasm. Additionally, the strain is deficient for the Ion and OmpT proteases, making it more amenable for protein production by heterologous gene expression.

Genotype: *fhuA2 lacC::T7 gene1 [Ion] ompT ahpC gal λatt::pNEB3-r1-cDsbC(*Spec^{R}*, lacI^{q}) ΔtrxB sulA11 R(mcr-73::mini*Tn*10*–Tet^{S}*)2 [dcm] R(zgb-210::*Tn*10*–TetS*) endA1 Δgor Δ(mcrC-mrr)114::*IS*10*

5.3.2 *S. cerevisiae* strains

BY4741

Laboratory strain, which is derived from S288c and was used for EUROFAN, the international systematic *S. cerevisiae* gene disruption project (Winzeler et al. 1999). It carries a mutation in the Flo8 transcription factor responsible for the expression of the flucculins *FLO1* and *FLO11* causing nonadhesive phenotypes (Bester et al. 2006; Fichtner et al. 2007). In this work the strain was used for adhesion assays on human epithelial cells. Therefore, it was transformed with plasmids carrying different *EPAA* domains fused to an *FLO11BC* domain. The BC domain of the flocculin Flo11 here serves as a platform to ensure a proper presentation of the EpaA domain on the cell surface. As the expression of flocculins in the S288c strain background is inhibited, the plasmid based expression system uses the *PGK1* promoter.

Genotype: *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*

5.3.3 Human cell lines

Human epithelial cell lines were used for *in vivo* adhesion assays. The cell lines were obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures). All cell lines are derived from tumor cells isolated from human patients and are classified as biosafety level 1.

Caco-2 (ACC 169)

The human line Caco-2 was originally isolated from a human epithelial colorectal adenocarcinoma in 1974 and has since been used as a laboratory model for the human intestinal epithelium, cell differentiation and viral transfection. Ploidy of Caco-2 cells can range from diploid to tetraploid dependent on cultivation time. Caco-2 has also been used to study the host cell adhesion of several *Candida* species (Dieterich et al. 2002; Gil et al. 2010; Negri et al. 2011; Maestre-Reyna et al. 2012; Diderrich et al. 2015). Here, the cell line was used for adhesion tests with EpaA variants of *C. glabrata* presented by *S. cerevisiae*.

HeLa (ACC 57)

The HeLa cell line was established from an epithelial cervix carcinoma in 1951 providing the first human cells that could be continuously maintained under laboratory conditions. HeLa cells

were extensively used for virus infection studies, tumor research and the development of tools for molecular diagnostics in human medical treatment. HeLa cells also form a confluent layer of cells but also tend to form cellular aggregates when reaching higher cell densities. This cell line was chosen because vaginal infections are among the most frequent infections with *C. glabrata*.

TCC-SUP (ACC 377)

Cell line isolated from an urinary bladder cell carcinoma in 1974 which grows as a monolayer of epithelial-like adherent cells. This cell line was chosen because it provides host cells different from the commonly used Caco-2 cell line and also represents tissue from another target organ. Generally, TCC-SUP cells form a monolayer under culture conditions and therefore can be used as a substrate for host cell adhesion tests.

5.3.4 Plasmids for expression of epithelial adhesion domains in *E. coli* and *S. cerevisiae*

In this work a number of plasmids have been constructed to express various natural and mutant *EPAA* variants in *E. coli* and *S. cerevisiae*. Plasmids for *E. coli* were based on the pET-28(a)+ expression system (*Merck*) which provides an N-terminal 6×His tag and is under the control of the T7 expression system. Expression was induced by addition of IPTG to a final concentration of 10 µM For expression in *S. cerevisiae* a YCplac33 based expression system was used that contained the promoter, signal sequence, BC domain and terminator of the *FLO11* adhesin. Different natural or mutant EpaA variants were inserted between the 3HA and BC domain via restriction cloning. A list of all plasmids used in this work is given in Table 14.

Table 14: Plasmids used for protein production and *in vivo* analyses.

Plasmid	Relevant genotype	Source
pET-28(a)+	P_{T7} *6xHis lacI* Kan^R	Merck, Germany
BHUM1829	$EPA1^{31\text{-}271}$ in pET-28(a)+	Diderrich et al. 2015
BHUM1790	$EPA6^{26\text{-}271}$ in pET-28(a)+	Diderrich et al. 2015
BHUM1886	$EPA9^{39\text{-}305}$ in pET-28(a)+	Diderrich et al. 2015
BHUM2495	$EPA10^{39\text{-}306}$ in pET-28(a)+	Diderrich et al. 2015
BHUM1873	$EPA12^{26\text{-}292}$ in pET-28(a)+	Diderrich et al. 2015
BHUM1875	$EPA15^{37\text{-}277}$ in pET-28(a)+	Diderrich et al. 2015
BHUM1876	$EPA23^{22\text{-}279}$ in pET-28(a)+	Diderrich et al. 2015
BHUM2498	$EPA1^{31\text{-}271;E227D;D229H}$ in pET-28(a)+	Diderrich et al. 2015
BHUM2510	$EPA9^{39\text{-}305;D258E;H260D}$ in pET-28(a)+	Diderrich et al. 2015
BHUM3157	$EPA1^{31\text{-}271;R71\text{-}P77::L73\text{-}Q108(EPA9)}$ in pET-28(a)+	This study
BHUM3158	$EPA6^{26\text{-}271;L71\text{-}P77::L73\text{-}Q108(EPA9)}$ in pET-28(a)+	This study
BHUM3159	$EPA9^{39\text{-}305;L73\text{-}Q108::R71\text{-}P77(EPA1)}$ in pET-28(a)+	This study

Continuation on next page.

Plasmid	Relevant genotype	Source
BHUM3160	*EPA9*$^{39\text{-}305;L73\text{-}Q108::L73\text{-}Q109(EPA10)}$ in pET-28(a)+	This study
BHUM3161	*EPA10*$^{39\text{-}306;L73\text{-}Q109::L73\text{-}Q108(EPA9)}$ in pET-28(a)+	This study
BHUM3189	*EPA1*$^{31\text{-}271;E227D;D229H;R71\text{-}P77::L73\text{-}Q108(EPA9)}$ in pET-28(a)+	This study
BHUM3190	*EPA9*$^{39\text{-}305;D258E;H260D;L73\text{-}Q108::R71\text{-}P77(EPA1)}$ in pET-28(a)+	This study
BHUM3381	*EPA12*$^{26\text{-}292;Y227F}$ in pET-28(a)+	This study
BHUM3384	*EPA15*$^{37\text{-}277;Y227F}$ in pET-28(a)+	This study
BHUM3388	*EPA23*$^{22\text{-}279;K231A}$ in pET-28(a)+	This study
BHUM1760	P_{FLO11}-*FLO11*$^{(1\text{-}30)}$-*SP-3HA-FLO11*$^{(214\text{-}1360)}$-T_{FLO11} in YCplac33	Diderrich et al. 2015
BHUM1983	*EPA1*$^{31\text{-}271}$ in BHUM1760	Diderrich et al. 2015
BHUM2018	*EPA6*$^{26\text{-}27}$ in BHUM1760	Diderrich et al. 2015
BHUM2020	*EPA9*$^{39\text{-}305}$ in BHUM1760	Diderrich et al. 2015
BHUM3167	*EPA10*$^{39\text{-}306}$ in BHUM1760	This study
BHUM2153	*EPA12*$^{26\text{-}292}$ in BHUM1760	Diderrich et al. 2015
BHUM2154	*EPA15*$^{37\text{-}277}$ in BHUM1760	Diderrich et al. 2015
BHUM3379	*EPA22*$^{28\text{-}266}$ in BHUM1760	This study
BHUM1988	*EPA23*$^{22\text{-}279}$ in BHUM1760	Diderrich et al. 2015
BHUM2160	*EPA1*$^{31\text{-}271;E227D;D229H}$ in BHUM1760	This study
BHUM2194	*EPA9*$^{39\text{-}305;D258E;H260D}$ in BHUM1760	This study
BHUM3192	*EPA10*$^{39\text{-}306;D259E;H261D}$ in BHUM1760	This study
BHUM3162	*EPA1*$^{31\text{-}271;R71\text{-}P77::L73\text{-}Q108(EPA9)}$ in BHUM1760	This study
BHUM3163	*EPA6*$^{26\text{-}271;L71\text{-}P77::L73::Q108(EPA9)}$ in BHUM1760	This study
BHUM3165	*EPA9*$^{39\text{-}305;L73\text{-}Q108::R71\text{-}P77(EPA1)}$ in BHUM1760	This study
BHUM3166	*EPA9*$^{39\text{-}305;L73\text{-}Q108::L73\text{-}Q109(EPA10)}$ in BHUM1760	This study
BHUM3191	*EPA10*$^{39\text{-}306;L73\text{-}Q109::R71\text{-}P77(EPA1)}$ in BHUM1760	This study
BHUM3164	*EPA10*$^{39\text{-}306;L73\text{-}Q109::L73\text{-}Q108(EPA9)}$ in BHUM1760	This study
BHUM3185	*EPA1*$^{31\text{-}271;E227D;D229H;R71\text{-}P77::L73\text{-}Q108(EPA9)}$ in BHUM1760	This study
BHUM3187	*EPA9*$^{39\text{-}305;D258E;H260D;L73\text{-}Q108::R71\text{-}P77(EPA1)}$ in BHUM1760	This study
BHUM3193	*EPA10*$^{D259E;H261D;L73\text{-}Q109::R71\text{-}P77(EPA1)}$ in BHUM1760	This study
BHUM3382	*EPA12*$^{26\text{-}292;Y227F}$ in BHUM1760	This study
BHUM3385	*EPA15*$^{37\text{-}277;Y227F}$ in BHUM1760	This study
BHUM3389	*EPA23*$^{22\text{-}279;K231A}$ in BHUM1760	This study

6 Methods

6.1 Cultivation of microorganisms

6.1.1 Cultivation of *Escherichia coli*

E. coli cultures were grown at 37 °C in LB rich medium supplemented with either 100 µM Ampicillin or 35 µM Kanamycin. Precultures were incubated overnight at 37 °C and used to inoculate main cultures at a dilution of 1:100. Cell growth was monitored by photometric measurements at 600 nm. An optical density of 1 OD_{600} corresponds to $2 \cdot 10^8$ cells per ml (Sherman et al. 1987). For heterologous protein production in *E. coli* the cultures were first grown at 37 °C and then shifted to 12 °C after induction of gene expression. For glycerol stocks the cells were grown overnight in the appropriate medium and supplemented with 30 % glycerol. Afterwards they were stored at −70 °C.

6.1.2 Cultivation of *Saccharomyces cerevisiae*

Cultivation of *S. cerevisiae* was performed at 30 °C in YPD, or in SC-Ura medium for auxotrophic selection on plasmids. For epithelial cell adhesion the cells were grown LF medium (LFM) to minimize autofluorescence while also using the same auxotrophic selection like in SC-Ura. During growth in liquid media the cell growth was monitored by spectroscopic measurement at 600 nm. An OD_{600} = 1 corresponds to $1.5 \cdot 10^7$ yeast cells per ml (Sherman 1991). Glycerol stocks were prepared by scraping cells from solid media, resuspending them in 30 % glycerol and storing them at −70 °C.

6.2 Cultivation of human cells

6.2.1 Propagation of human cells by culture splitting

In this work the three human epithelial cell lines Caco-2 (colorectal adenocarcinoma), HeLa (cervix adenocarcinoma) and TCC-SUP (urinary bladder transitional cell carcinoma) were used. All cell lines were cultured at 37 °C and 5 % CO_2 in 10 ml Dulbecco's Modified Eagle Medium (Thermo Fisher Scientific). The medium was supplemented with 15 % fetal bovine serum (FBS) for Caco-2 or 10 % FBS for HeLa and TCC-SUP cells, according to the recommended value for each cell line. Additionally 1 % Pen/Strep was added to all DMEM media. Culture passaging was performed upon reaching a confluence of >80 %. The first step in subculturing was the dissociation of adherent cells. Therefore, the growth medium was removed and the cells were washed once with 10 ml PBS. Then 1 ml trypsin was added and the cells were incubated for 10 min at 37 °C and 5 % CO_2. Afterwards, 9 ml DMEM + FBS was added to stop the enzymatic reaction. This suspension was used to inoculate a new culture in a 1:10 ratio.

6.2.2 Storage of human cells

For the storage of human cell lines the cells were treated with trypsin in the same way as for subculturing. After the addition of 9 ml DMEM with 10 to 15 % FBS and 1 % Pen/Strep, the cells were centrifuged at 300 × g for 5 min. The pellet were resuspended in 1 ml ice cold FBS with 10 % DMSO and immediately stored at −80 °C. After one week the vials were transferred into a storage container with liquid nitrogen for long term storage.

6.2.3 Mycoplasm testing

All cell lines were regularly tested for mycoplasm contamination by PCR. Therefore the kit "Venor GeM One Step" was used. After amplification, the PCR products were separated on a 1.8 % agarose gel. Detection of DNA bands was performed under UV light at 254 nm with a GelDoc XR System and the Quantity One 1D software (Bio-Rad).

6.3 Transformation of *E. coli* and *S. cerevisiae*

6.3.1 Preparation of chemical competent *E. coli*

For the preparation of chemical competent cells 5 ml LB were inoculated with the appropriate *E. coli* strain and incubated over night at 37 °C. On the next day 50 ml LB medium were inoculated with 0.5 ml of the overnight culture. This main culture was grown at 37 °C until reaching an OD_{600} of 0.4 to 0.6. The cells were then cooled to 4 °C, harvested by centrifugation at 4000 × g for 15 min and resuspended in 16 ml RF1 buffer. The suspension was incubated on ice for 15 min and centrifuged again at 4 °C with 4000 × g for 15 min. Afterwards, the cells were resuspended in 4 ml RF2 buffer and incubated for 15 min on ice. Finally, the cell suspension was split into 100 µl aliquots, frozen in liquid nitrogen and stored at −80 °C.

6.3.2 Transformation of chemical competent *E. coli*

One aliquot of chemical competent *E. coli* cells was thawed on ice for each transformation and 1 to 5 µl of a ligation or 0.2 to 1 µl of a plasmid preparation (containing 1 pg–100 ng DNA) was added and mixed by carefully flicking the tubes. The reactions were incubated for 30 min on ice and then heat shocked for 30 s at 42 °C. After cooling down the cells on ice for 5 min, 900 µl LB medium was added. Then the mixture was placed at 37 °C for 60 min before spreading the cells on selection plates with appropriate antibiotics.

6.3.3 Preparation of competent *S. cerevisiae* (Knop et al. 1999, modified)

The BY4741 yeast strain was grown overnight at 30 °C in 10 ml YPD medium. On the next day 50 ml YPD medium were inoculated with the preculture and grown from an OD_{600} of 0.2 to 0.6 at 30 °C. Then the cells were harvested by centrifugation at 850 × g for 3 min. Subsequently the cells were washed twice with 15 ml sterile water and once with 10 ml SORB buffer. The resulting

pellet was resuspended in 360 µl SORB buffer and 40 µl freshly denatured salmon sperm DNA was added and the mixture was divided into 50 µl aliquots. This aliquots could either be stored at −80 °C or directly be used for transformation.

6.3.4 Transformation of competent *S. cerevisiae*

Competent *S. cerevisiae* cells were thawed on ice for each transformation and 1 to 2 µg plasmid was added together with 600 µl PEG buffer per aliquot. After incubation for 30 min at room temperature the cells were heat shocked at 42 °C for 15 min. Then the cells were centrifuged (3 min, 380 × g), washed once with 1 ml YPD medium and resuspended in 1 ml YPD. After regenerating of 60 min, the cells were spread on a selection plate and incubated for three days.

6.4 Preparation and manipulation of DNA

6.4.1 Isolation of plasmids from *E. coli*

A single colony of an *E. coli* was inoculated in 5 ml LB with ampicillin or kanamycin and grown overnight at 37 °C. The cells were then harvested by centrifugation for 1 min at 16 000 × g. The plasmid DNA was isolated via the "NucleoSpin Plasmid kit" (Macherey-Nagel GmbH & Co. KG) or the "E.Z.N.A. Plasmid DNA Mini Kit" (Omega Bio-tek, Inc) according to the manufacturer's instructions. The DNA concentration was measured photometrically with a "NanoDrop ND-1000" (Thermo Fisher Scientific) and stored at −20 °C. For photometric determination of DNA concentration, the absorption of a given sample at a wavelength of 260 nm is measured. To additionally determine the purity of the sample, the protein content is also determined by absorption measurement at 280 nm. DNA concentration can then be calculated by formula [1]:

$$c = A_{260} \cdot 50\,\mathrm{ng/\mu l} \qquad [1]$$

A: absorbance, c: DNA concentration

6.4.2 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify and modify DNA fragments. Therefore, specific oligonucleotides (primers) were designed, which are complementary to the 3' end (forward primer) or the 5' end (reverse primer). Primers usually had a length of 18 to 25 nucleotides, but for the introduction of mutations, longer primers with up to 53 nt were also used. For the reactions the "Phusion High-Fidelity DNA Polymerase" (Thermo Fisher Scientific), together with the recommended HF buffer was used. The reaction mix (50 µl) and temperature profile are given below (Table 15 and Table 16). The temperature profile was adjusted adjusted for each reaction according to the manufacturer's instructions to match the annealing temperature of the respective primers. PCR products were checked by agarose gel electrophoresis and purified

by using either the "E.Z.N.A. Gel Extraction Kit" (PCR purification protocol) or by ethanol precipitation. Therefore 100 µl ice cold 100 % ethanol were added and the reaction tube was stored at −20 °C for 10 to 60 min. Then the tube was centrifuged at 16 000 × g for 5 min and the supernatant was discarded. The pellet was washed once with 70 % ethanol, dried and resuspended in 30 µl sterile water.

Table 15: PCR reaction mix

Component	Volume (µl)	Final concentration
5 × HF buffer	10	1 ×
10 mM dNTPs	1	200 µM
Forward primer	x	0,5 µM
Reverse primer	x	0,5 µM
Template DNA	x	
Phusion DNA polymerase	0.5	0,02 U/µl
H_2O	ad 50	

Table 16: PCR temperature profile

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	98	30	
Denaturation	98	5 to 10	25 ×
Annealing	X	10 to 30	
Extension	72	15 to 30 per kb	
Final Extension	72	300	

6.4.3 Site directed mutagenesis (Weiner et al. 1994)

The introduction of mutations into plasmid DNA was carried out via site directed mutagenesis, a PCR-based method to introduced mutations at a specific location in the target DNA. Therefore, it uses complementary primers containing the desired mutations to modify the template DNA in a whole plasmid amplification reaction. Each primer was designed with 15 nt complementary sequence at each side (3' and 5') of the target site. The linear amplification product was digested with DpnI at 37 °C to specifically remove methylated template DNA. In a subsequent reaction, the linear PCR product was ligated with T4 DNA ligase. Afterwards, the ligation reaction was used for direct transformation of chemically competent Top10 *E. coli*. For verification the plasmid DNA was isolated from *E. coli* and tested by sequencing for the previously introduced mutations.

6.4.4 Restriction of DNA

Analytical restriction was used for restriction fragment length analysis to test plasmids for the insertion of DNA fragments. Therefore the plasmid was digested with appropriate restriction enzymes and separated on an 1 % agarose gel to check for fragment length differences. For preparative DNA restrictions up to 5 µg DNA was digested and separated on an 1 % agarose gel. Bands of the right length were cut out and the DNA was isolated using the E.Z.N.A. Gel Extraction Kit (Omega Bio-tek, Inc.).

6.4.5 Ligation of DNA fragments

Linear DNA fragments were covalently connected by ligation via T4 DNA ligase (Roche Applied Science). Therefore the DNA fragments generated by PCR or DNA restriction were incubated in ligation buffer with 1 to 5 U ligase for at least one hour at room temperature. Depending on fragment size and overhang length at the end of the fragments, a ratio of 1:1 to 5:1 was used, according to manufacturer's instructions. The volume for all ligation reactions was 20 µl.

6.4.6 Agarose gel electrophoresis

Separation of DNA fragments was performed by electrophoresis on 1 % agarose gels in TAE buffer. The samples were mixed with 6 × DNA loading dye, loaded to the gel and run at 100 to 120 V for 30 to 60 min. For reference, the "Generuler DNA ladder mix" was used. After electrophoresis, the gel was stained for 20 min in a GelRed solution and subsequently destained for 10 min in distilled H_2O. DNA bands were visualized and documented with a GelDoc XR System and the Quantity One 1-D Software (BioRad).

6.5 Biochemical methods

6.5.1 Heterologous production of proteins in *E. coli*

For the production of soluble protein all *EPAA* variants were heterologously expressed in the *E. coli* SHuffle T7 Express strain using an pET-28a(+) plasmid system. Each construct was tested first for protein production at analytical scale. Therefore, a 5 ml preculture in LB+Kan was grown overnight at 37 °C and used to inoculate a 50 ml LB+Kan main culture at a ratio of 1:100. This main culture was grown up to a OD_{600} of 0.6 and then inoculated with 0.05 mM IPTG to induce *EPAA* expression and shifted to 4 °C. The culture was then incubated for 72 h at 4 °C and subsequently harvested by centrifugation (4000 × g, 15 min). Afterwards, the pellet was resuspended in 500 µl AM+ buffer and transferred to screw cap reaction tubes containing glass beads of 0.2 to 0.25 mm diameter for cell lysis. Then, 10 µl lysozyme was added and the cells were incubated for 30 min at room temperature followed by cell lysis in the MP Bio FastPrep 24 homogenizer (Kem-En-Tec Nordic). A sample of the supernatant (10 µl) was mixed with 10 µl HU-buffer and denatured at 65 °C for 10 min. Additionally, a sample of the pellet was

prepared. Therefore, it was resuspended in 500 µl AM+ buffer and a 10 µl sample was taken. This sample also was mixed with 10 µl HU-buffer and denatured at 60 °C for 10 min. Samples were then analyzed by SDS PAGE and Coomassie Brilliant Blue staining for comparison of protein production levels.

Preparative protein production of EpaA variants was performed in the same way as the production on analytical scale. Therefore, an *E. coli* preculture of 50 ml was grown overnight in LB+Kan at 37 °C. On the next day, the main culture was inoculated at a ratio of 1:100 and grown at 37 °C to an OD_{600} of 0.6 to 0.8 while shaking at 120 rpm. Afterwards, the temperature was reduced to 12 °C and protein production was induced by addition of 0.05 mM IPTG. The cultures were harvested after incubation for 72 h by centrifugation at 4 °C (8000 × g, 16 min). The pellet was resuspended in 15 ml AML buffer, shock frosted in liquid nitrogen and stored at −80 °C.

6.5.2 Purification of proteins produced in *E. coli*

Purification of proteins was performed by lysis of *E. coli* cells followed by a combination of immobilized metal ion affinity chromatography (IMAC) and size exclusion chromatography (SEC). The pET-28a(+) plasmid system provides two hexa-histidin tags, one for N-terminal and one for C-terminal tagging. In this work all proteins carried an N-terminal 6 ×His tag for affinity chromatography using columns packed with nitrilotriacetic acid and Ni^{2+} ions (Ni-NTA). In a first step, the *E. coli* cells were lysed by using a French press. In this method the cells are subjected to high pressure and then passed through a small diameter valve. The shear forces that occur in this process cause the cell walls to rupture and the cells are lysed. Here, frozen *E. coli* cells were prepared for lysis by thawing and addition of 10 µl EDTA (0.5 M), 80 µl lysozyme (50 mg/ml), 50 µl PMSF (1 M) and one spatula tip of deoxyribonuclease I (DNAseI). Afterwards, the suspension was incubated for 30 min at 4 °C and then transferred to a pre-cooled pressure cell. Lysis of *E. coli* cells then was performed by three passes at a pressure of about 7 MPa. Insoluble fragments were separated by centrifugation (4 °C, 45 min, 111 111 × g) and the supernatant was filtered with a syringe filter with a pore size of 0.45 µm (Merck Millipore). Then, the filtrate was transferred to a 2 ml Protino Ni-NTA column (Macherey-Nagel) by using a P1 peristaltic pump (Pharmacia Biotech). The flow-through was discarded and the column was subsequently washed with AM buffer containing 10 mM and 20 mM imidazole to remove nonspecifically bound proteins. Afterwards, the 6×His fusion protein was eluted with 10 ml AMI buffer (250 mM imidazole) and the column was washed with each 20 ml AMI buffer (500 mM imidazole) and distilled H_2O, respectively.

In a next step, the eluted protein solution was concentrated to 1 ml and filtered using Amicon Centrifugal Filter Devices (Merck Millipore). The filtrate was further purified by SEC using a HiLoad Superdex 75 prep grade column (*GE Healthcare*) to separate the proteins by their apparent size. Therefore, the column was first equilibrated with 120 ml AML buffer and then loaded with the protein solution. All steps were performed using an ÄKTA Purifier (*GE Healthcare*) and

a buffer flow rate of 1 ml/ml. The flow-through was fractionated into 1 ml aliquots and tested by SDS PAGE for purity. Protein containing fractions were pooled and finally concentrated for further analysis.

6.5.3 Buffer exchange

For a detailed analysis of purified EpaA variants, different buffer systems were used. Buffer exchange was performed by using PD10 desalting columns (*GE Healthcare*) with a gravity flow protocol. Here, the column was first equilibrated with the target buffer before applying a protein aliquot of up to 2 ml. Then, target buffer was added to elute the protein in fractions of 500 µl. Protein concentration of each fraction was tested by using a NanoDrop ND-1000 spectrometer (*Thermo Fisher Scientific*). Protein containing fractions were pooled and concentrated to 1 ml.

6.6 Analytical methods

6.6.1 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is an analytical method to separate proteins on a gel by their apparent mass inside an electric field (Laemmli 1970). Therefore, a protein sample is denatured by heating and addition of high concentrations of urea and SDS (HU buffer). Moreover, dithiothreitol (DTT) is added to reduce disulfide bonds and completely unfold the protein. During denaturation, the charge of each protein gets covered by SDS resulting in proteins with an overall negative charge. Electrophoresis then usually is performed on gels with two layers containing different concentrations of acrylamide for a better resolution of protein bands (Table 17). Here, a discontinuous SDS-PAGE with acrylamide concentrations of 4 % and 12 % were used. Proteins were denatured at 65 °C before loading to the gel. Electrophoresis was performed at 150 V until the buffer front reached the lower end of the gel. The PageRuler Prestained Protein Ladder (*Thermo Fisher Scientific*) was used as a mass standard.

Table 17: Composition of polyacrylamide gels (2 gels)

Component	Separating gel (12 %) (ml)	Loading gel (4 %) (ml)
Tris pH 8.8 (1 M)	3.75	
Tris pH 6.8 (0.5 M)		2.5
Acrylamide (30 %)	2.7	1.3
SDS (10 % (w/v))	0.1	0.1
APS (10 % (w/v))	0.1	0.1
TEMED	0.005	0.005
H_2O	1.6	6.1

After electrophoresis, protein bands in the separating gel were visualized by Coomassie Brilliant Blue staining. Therefore, the gel was first covered in staining solution for 30 min and then destained overnight.

6.6.2 Determination of protein concentration

The concentration of protein domains purified from *E. coli* was determined photometrically. Therefore, the molar extinction coefficient (ε_{280}) of a given protein sequence was first determined using the ProtParam server (Gasteiger et al. 2005). The coefficient is always given for a wavelength of 280 nm since light absorption by proteins mainly depends on the presence of aromatic amino acids like tyrosin and tryptophan. Additionally, the presence of disulfide bonds was also taken into account, since these also affect light absorption. Thus, the protein concentration of a solution can be determined by Lambert–Beer law [2]:

$$A = \varepsilon_{280} \cdot c \cdot l \qquad [2]$$

A: absorbance, ε_{280}: molar extinction coefficient (280 nm), c: molar concentration, l: optical path length, m: mass, V: volume, M: molecular weight

Table 18: Molar extinction coefficients and molecular weights of EpaA variants

Variant	Identifier in strain collection	Molar extinction coefficient (l/(mol · cm))	Molecular weight including 6 × His-tag (g/mol)
Epa1A	BHUM1829	41 175	29 401.63
Epa6A	BHUM1790	48 165	29 909.01
Epa9A	BHUM1886	54 125	33 203.52
Epa10A	BHUM2495	49 655	32 951.15
Epa12A	BHUM1873	43 695	32 995.16
Epa15A	BHUM1875	47 135	29 960.77
Epa22A	BHUM2496	37 165	28 959.63
Epa23A	BHUM1876	41 175	31 765.25
Epa1A^{CBL2Epa2}	BHUM1804	42 665	33 076.85
Epa1A^{CBL2Epa3}	BHUM1805	39 685	29 251.52
Epa1A^{CBL2Epa6}	BHUM1806	42 665	33 077.83
Epa1A^{CBL2Epa9}	BHUM2498	41 175	29 409.65
Epa9A^{CBL2Epa1}	BHUM2510	54 125	33 195.50
Epa1A^{L1Epa9}	BHUM3157	40 800	32 254.42
Epa6A^{L1Epa9}	BHUM3158	49 655	32 894.99
Epa9A^{L1Epa1}	BHUM3160	54 125	30 350.73

Continuation on next page.

Variant	Identifier in strain collection	Molar extinction coefficient (l/(mol · cm))	Molecular weight including 6 × His-tag (g/mol)
Epa9A^{L1Epa10}	BHUM3161	54 125	33 294.69
Epa10A^{L1Epa9}	BHUM3159	49 655	32 859.98
Epa1A$^{CBL2+L1Epa9}$	BHUM3189	41 175	32 262.44
Epa9A$^{CBL2+L1Epa1}$	BHUM3190	54 125	30 342.71
Epa12A^{Y227F}	BHUM3381	42 205	32 979.16
Epa15A^{Y227F}	BHUM3384	45 645	29 944.77
Epa23A^{K231A}	BHUM3388	41 175	31 708.16

6.6.3 Fluorescent labeling of proteins

For glycan array analysis the EpaA proteins were labeled with Alexa Fluor 488 TFP ester (*Thermo Fisher Scientific*). Protein labeling with Alexa Fluor 488 TFP ester covalently links the fluorescent molecule to primary amines of the target molecule. In this work, all EpaA variants were prepared for labeling by buffer exchange to prevent reaction of the TFP ester with the Tris contained in the AML buffer used for protein purification. To ensure suitable conditions for the labeling reaction, a bicarbonate buffer with pH 8.3 was used. For each reaction, 1 ml of the respective protein solution (10 mg/ml) was then mixed with 50 µl of the amine-reactive dye, according to manufacturer instructions. The reaction mix was then incubated for 90 min before separating unbound dye by gel filtration using PD10 desalting columns (*GE Healthcare*). Afterwards, another buffer exchange was performed to transfer the proteins back to AML buffer. For storage, the labeled proteins were kept at 4 °C to avoid degradation.

6.6.4 Glycan array analysis

Glycan array analysis was performed to determine ligand binding specificity of EpaA lectins produced by heterologous gene expression in *E. coli*. Array experiments were conducted by the Consortium of Functional Glycomics (CFG) with Alexa Fluor 488 labeled proteins. The chips used for EpaA analysis were Mammalian Screen V 5.4 arrays containing 585 different glycan structures. These glycans differ in length, composition and complexity to represent a large variety of carbohydrate structures present on human cell surfaces. Purified EpaA proteins were prepared by fluorescent labeling and concentration to 1 mg/ml and then sent to the CFG for glycan array analysis. There, the proteins were diluted to 5 µg ml^{-1} or 50 µg/ml, respectively, and applied to the chip. After subsequent washing the amount of bound protein was quantified by fluerescence measurement. For evaluation of the data all fluorescence levels were converted to relative values by assigning a value of 1 to the best bound carbohydrate on the chip. A detailed description of the protocol was given by Heimburg-Molinaro et al. (2011).

6.6.5 Fluorescence titration spectroscopy

Fluorescence titration spectroscopy can be used for the characterization of protein-ligand interactions. This is based on the principle that proteins that are excited with ultraviolet light(λ = 280 nm) can emit fluorescent light. Ligand binding results in a reduced emission (quench) of fluorescent light, which can be used to calculate binding constants. The different energy states and effects involved in this process can be visualized in an Perrin-Jablonski diagram (Figure 34). It is named after the two physicists Jean Perrin and Aleksander Jablonski who first described the processes visualized in the diagram. For the analysis of protein-ligand interaction either the intrinsic fluorescence of aromatic amino acids or a fluorescent label can be used.

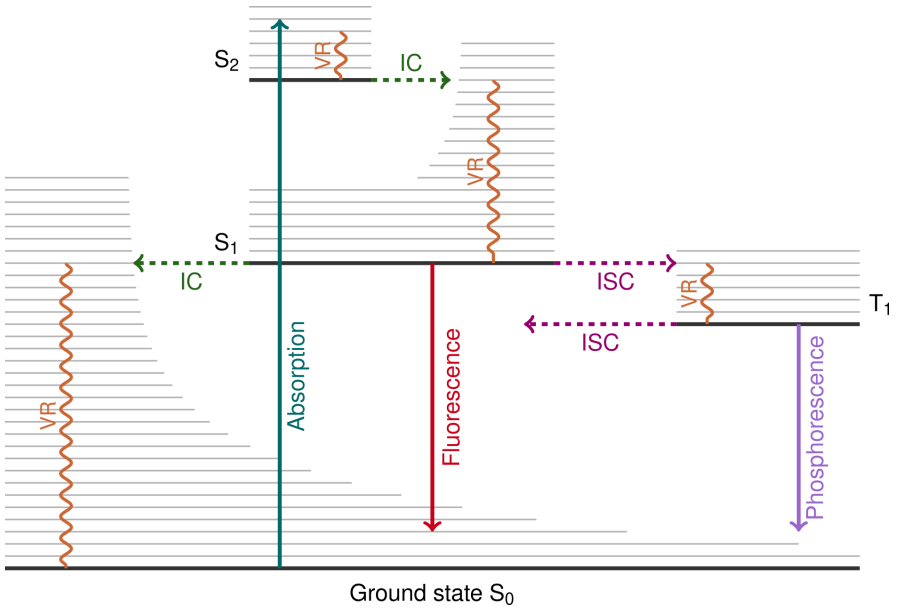

Figure 34: Perrin-Jablonski diagram. By the absorption of a photon (green) the molecule's electrical state changes from a ground state (S_0) to an excited state (S_2). Nonradiative deactivation by internal conversion (IC) and vibrational relaxation (VR) causes the molecule to enter a lower energy state (S_1). From the S_1 state the molecule can again deactivate via IC and VR and return to S_0. Alternatively, the transition from S_1 to S_0 can also occur by the emission fluorescent light (red). As a third option, the molecule may enter an intermediary triplet state (T_1) by intersystem crossing (ISC) and VR and finally deactivate by the emission of phosphorescencent light (purple). This figure is a derivative of "The Perrin-Jablonski diagram" by Germain Salvato-Vallverdu (Salvato-Vallerdu 2009), used under CC BY 2.5.

The energy levels of a molecule are called electric states (S_0-S_2) and can each have several vibrational states. Upon absorption of a photon, the molecule's energy level changes from the ground electric state (S_0) to an excited state (S_2). Returning from this excited state back to the ground state can be caused by different types of radiative or nonradiative deactivation. The term radiative deactivation describes the emission of a photon and comprises fluorescence ($S_1 \rightarrow S_0$) and phosphorescence ($T_1 \rightarrow S_0$). Nonradiative deactivation occurs via internal conversion (IC) and vibrational relaxation which can take place from either vibrational states or excited electric states. Another effect, called intersystem crossing (ISC), can cause the molecule to switch between a higher energy singlet state S_1 and the lower energy triplet state T_1. In this work, fluorescence titration spectroscopy was used for binding studies of different EpaA variants to Galβ1-3GalNAc

(T-antigen), galactose and 6SGal (sulfogalactose). Therefore, heterologously produced proteins were used that were purified by affinity chromatography and size exclusion chromatography. Since the purified EpaA domains contained residual lactose from the purification process, 10 µl EDTA was added to remove any Ca^{2+} ions and release the bound lactose. Then, a buffer exchange was performed by using PD10 desalting columns (*GE Healthcare*) to remove the unbound lactose. Finally, the EpaA domains were titrated with the different ligands while the emitted fluorescent light was measured. Ligand concentrations and volumes used for titration are shown in Table 19.

Table 19: Ligand volume for fluorecence titration

Injection	Stock concentration (mM)	Ligand volume (ml)	Final concentration (µM)
1	0	0	0.00
2	0.125	1	0.12
3	0.125	1	0.25
4	0.125	2	0.50
5	0.125	4	0.99
6	0.125	8	1.97
7	1	2	3.93
8	1	2	5.88
9	1	4	9.77
10	1	4	13.62
11	1	8	21.24
12	1	8	28.74
13	1	8	36.12
14	10	2	55.03
15	10	3	83.25
16	10	4	120.64
17	10	4	157.75
18	10	6	212.89

6.7 Determination of protein structure

The determination of three-dimensional protein structures is a powerful tool to understand the structure-function relationship of a protein and its involvement in biological processes. It not only provides valuable information about the spatial arrangement of a known protein sequence, but can also uncover protein-protein or protein-ligand interactions. Today, a variety of different methods is available for protein structure determination. These are X-ray crystallography (Kendrew et al. 1958), small-angle X-ray scattering (SAXS) (Lipfert and Doniach 2007), nuclear magnetic

resonance spectroscopy (NMR) (Wüthrich 2001) and cryogenic electron microscopy (Cryo-EM) (Kühlbrandt 2014). Cryo-EM is mostly used for the analysis of quaternary protein structures like large protein complexes, while SAXS and NMR are used for the resolution of medium sized and small protein structures, respectively. According to the worldwide Protein Data Bank (www.wwpdb.org), the most commonly used method for resolving a protein structure at atomic resolution is X-ray crystallography followed by NMR (Berman et al. 2003). NMR has the advantage of visualizing flexible protein areas and also is able to work with proteins in solution which allows high throughput (Liu et al. 2005). Drawbacks of this method are that it is limited to proteins with a molar mass ≤30 kDa (Wüthrich 1990) and that proteins must be must be ^{13}C or ^{15}C spin-labeled before analysis. In contrast to that, X-ray crystallography is principally not limited by the molar mass of proteins and needs no labeling prior to structural analysis.

6.7.1 Protein crystallization

For the determination of a protein structure by X-ray crystallography, the protein first needs to be purified and crystallized using appropriate conditions. Since the identification of suitable crystallization conditions is an empiric process and can not be derived from the primary sequence of a protein, large sparse matrix screens were developed for testing several buffer conditions in parallel. These screens typically are based on conditions that have been proven successful in prior crystallization studies (Jancarik and Kim 1991; Stevens et al. 2001). Usually, three different parameters can be varied to change crystallization conditions: the pH and ionic strength of the buffer and the concentration of precipitants or metal ions. In this work, a number of different commercially available sparse matrix screens were used to set up sitting drop vapor diffusion experiments (Table 20).

Table 20: Crystallization screens

Screen	***Source***	**Description**
$AmSO_4$	*Qiagen*	Systematic screen of different $AmSO_4$ conditions
Classics	*Qiagen*	Literature based conditions
JCSG Core I–IV	*Qiagen*	Optimized conditions by the JCSG (Joint Center for Structural Genomics)
JCSG+	*Qiagen*	Precursor of JCSG core screens screens
MBC II	*Qiagen*	Optimized conditions for membrane proteins
Morpheus	*Molecular Dimensions*	PEG-based screen including ligands often found in the PDB (protein database)
Morpheus II	*Molecular Dimensions*	Includes heavy metals and a variety of other additives
PACT	*Qiagen*	PEG-based screen combined with different ions

Crystallization by vapor diffusion bases on the differences in osmolarity between a drop of crystallization mix and a buffer reservoir. This difference causes evaporation of water from the drop until its osmolarity equals the buffer solution. At the same time, protein concentration increases which leads to precipitation of protein and formation of crystal-nuclei. In this work, the EpaA variants were used in concentrations of 10 mg/ml or higher to obtain protein crystals. Crystallization mixes were prepared and pipetted to 96-well Innovadyne plates (*Agilent*) by using an automated pipetting robot (Cartesian Microsys SQ 4004, *Genomic Solutions*). Each plate contained one drop per buffer condition with 300 nl protein solution and 300 nl buffer and a second drop with half of the protein concentration. The plates were sealed with a transparent foil (VIEWseal, *Greiner Bio-One*) and kept at 18 °C. Documentation of crystal growth was performed using a Rock Imager system (*Formulatrix*). Photos of the crystallization plates were taken for the first three days at different time points (0 h, 6 h, 12 h, 24 h, 48 h and 72 h). Afterwards, the plates were checked regularly by using a microscope. Conditions of this initial screening that resulted in crystal growth were optimized to yield larger, more uniform crystals. Therefore multiple conditions similar to the initial screening were tested, using the hanging-drop method. The drops were prepared in 24-well plates (*Qiagen*) with 1 ml buffer reservoir using two different protein concentrations.

6.7.2 X-ray diffraction experiments

Crystals that grew to a sufficient size were picked with CryoLoops (*Hampton Research*) or MicroMounts (*MiTeGen*) and then were frozen in liquid nitrogen. Depending on the buffer conditions, 20 % glycerol was added as a cryoprotectant to avoid the formation of water crystals by the mother liquor. Reduction of the temperature to 100 K prevents the crystals from damage like the formation of radicals or disruption of disulfide bonds (Henderson 1990). At the beamline, the electron density of protein crystals was characterized by irradiation with X-rays and detection of the resulting diffraction patterns. This patterns depend on the symmetry and cell parameters of the crystal which is described by Bragg's law (Drenth and Mesters 2007). These data allow the calculation of a three-dimensional model of the protein structure at atomic resolution. The measurement was performed at the European Synchrotron Radiation Facility (ESRF) in Grenoble (France) and the BESSY II synchrotron in Berlin (Germany) by members of the workgroups of Prof. Dr. Lars-Oliver Essen and Prof. Dr. Gert Bange.

6.7.3 Processing and data reduction

Integration of diffraction data was performed with the XDS software (Kabsch 2010b). This takes all reflections measured in the experiment and applies corrections for beam-position, detector distance and mosaicity and also includes the exact cell parameters. At this point several parameters are used as criteria for data quality: the proportion of the measured reflexes to the theoretically possible reflexes (completeness), the difference in intensity between symmetry-equivalent reflexes (R_{merge}) and the signal-to-noise ratio ($I/\sigma(I)$). As a next step, the data scaled by merging

symmetry-equivalent reflections with XSCALE (Kabsch 2010a) or SCALA (Evans 2011), both run within the CCP4 software suite (Winn et al. 2011). The phase problem for Epa1A^{CBL2Epa9} was solved by molecular replacement with PHASER (McCoy 2007) using a carefully trimmed model of Epa1A (PDB 4AF9). For Epa9A^{CBL2Epa1} a lactose bound Epa9A (4CP0) was used as a model. The final refinement of both structures was done with alternating rounds of phenix.refine (Adams et al. 2010) and Coot (Emsley et al. 2010). The quality of the calculated protein structure was expressed using an R-factor comprising two values, R_{work} and R_{free}. While R_{work} is defined by the reflections used for refinement, the R_{free} value represents test-reflections that were selected during data reduction. Processing and refinement of data for Epa1A^{CBL2Epa9} (PDB 6Y9J) and Epa9A^{CBL2Epa1} (PDB 6Y98) were done by Viktoria Reithofer from the workgroup of Prof. Dr. Lars-Oliver Essen.

6.7.4 Preparation and visualization of protein structures

Figures of protein structures were prepared with the visualization tools PyMOL *Schrödinger* and UCSF Chimera . Modelling of EpaA domains was done with the MODELLER tool of UCSF Chimera and Epa1A (PDB 4A3X) and Epa9A (PDB 4CP0) as templates. Docking simulations for Epa9A, Epa12A and were performed with the AutoDock Vina utility of UCSF Chimera. Movies of molecular dynamics simulations were done with PyMOL

6.8 Functional analysis *in vivo*

6.8.1 Construction of *S. cerevisiae* expression plasmids

Functional analysis of Epa adhesion domains was done *in vivo* by using the BY4741 yeast strain of the EUROFAN collection. For expression of different *EPA* adhesion domains in this strain, a plasmid (YCplac33) based expression system under the control of the *PGK1*-promoter was used. This plasmid carried a *FLO11* secretion signal, a threefold hemagglutinin tag, the *FLO11BC* domain containing amino acids 214 to 1360 and the *FLO11* terminator. This *FLO11* expression cassette ensures the proper presentation of EpaA domains on the cell surface. Therefore, the adhesion domains were inserted between the signal peptide and the *FLO11BC* domain by restriction enzyme-based cloning using the endonucleases SacI and SacII.

6.8.2 Immunofluorescence microscopy

Yeast strains carrying *EPAA* expression plasmids were tested for expression and proper presentation of the respective adhesion domains by immunofluorescence microscopy. Therefore, the *S. cerevisiae* strains were grown overnight in LFM at 30 °C. At the next day, the cells were inoculated in fresh LFM at a rate of 1:100 and incubated for 4 h at 30 °C. For fluorescent labeling, 2 ml of this culture were first harvested by centrifugation at 500 × g and subsequently washed with 1 ml PBS+BSA. Then, the cells were incubated with the primary mouse anti-HA antibody (*Sigma aldrich*) for 1 h. After three additional wash steps, the secondary Cy3-labelled goat

anti-mouse antibody (Sigma Aldrich) was added and the cells were incubated for 30 min. Finally, the labelled yeast cells were again washed three times with PBS+BSA and then pipetted to agarose coated glass slides for microscopy. Relative fluorescence of each sample was determined by quantification of at least 100 cells for each EpaA variant. The fluorescence of the Epa1A presenting strain was set to 100 % and used as a reference.

6.8.3 Adhesion to mammalian cells

The A domains of epithelial adhesins have been shown to confer adhesion to human epithelial cells (Diderrich et al. 2015). In this work several mutant EpaA variants were created and tested for adhesion on human epithelial cells for comparison with the natural variants. To determine whether epithelial adhesins confer differential adhesion to epithelium of different origin, three cell lines were examined. For comparison with existent data, cells from human colon (Caco-2) were chosen as host cells for *in vivo* adhesion assays. Additionally, bladder (TCC-SUP) and cervix (HeLa) cells were also tested, since these tissues are also major infection sites for *C. glabrata*. Yeast cells carrying the appropriate expression plasmids were grown overnight in LFM-Ura at 30 °C.

On the next day, a main culture was inoculated at a dilution of 1:20 and incubated for 4 h. Then, 2 ODs of each culture were harvested by centrifugation (500 × g, 4 min) and washed twice with 500 µl PBS buffer containing 1 % BSA. In a next step, the cells were incubated with a primary rabbit anti-*Candida* antibody (*Acris*) diluted 1:1000 in 1 ml PBS+BSA for 60 min followed by three washing steps with PBS+BSA. Then, a secondary DyLight488-conjugated goat anti-rabbit antibody was added and the cells were incubated for 30 min in darkness. After three additional wash steps, 0.4 ODs of the labelled cells were added to a 48-well plate with human epithelial cells and incubated for 120 min. Host cells were prepared in 48-well plates by growing in DMEM (15 % FBS) until confluence. Before adding the labelled yeast cells, epithelial cells were washed with PBS and each well was filled with 125 µl DMEM (15 % FBS) without antibiotics. For the adhesion test, 125 µl of labelled yeast cells (OD_{600} = 0.4) were added to each well and incubated for 2 h at 37 °C in darkness.

After incubation, nonadhesive yeast cells were removed by carefully pipetting 250 µl PBS buffer to each well to avoid damaging the confluent epithelial cell layer. The supernatant was removed by inverting the plate followed by carefully tapping it on a paper towel. Then, all adhesive cells were removed by adding 200 µl PBS buffer and scratching with a pipet tip. The resulting suspension was transferred to a black 96-well plate for fluorescence measurement. Wells with only PBS, epithelial cells in PBS or labelled yeast cells in PBS (125 µl, OD_{600} = 0.4) were used as controls and for normalization against overall fluorescence. Additionally, an *EPA1A* expressing yeast strain was present on all plates as a reference. Fluorescence was measured with an Infinite 200 Pro fluorimeter (Tecan) using the settings in Table 21. For the evaluation of adhesion behavior, changes in adhesion between mutant and natural variants were tested

for statistical significance by an unpaired *t*-test. Here, *p*-values of <0.01 were considered as significant.

Table 21: Fluorimeter settings

Setting	Value
Excitation wavelength	488 nm
Excitation band width	9 nm
Excitation wavelength	520 nm
Excitation band width	20 nm
Gain	100
Excitation lights	25
Integration time	488 µs
Z-position	18 300 µm

6.9 Molecular dynamics simulation

Molecular dynamics simulation of L1 dynamics and protein-ligand interaction was performed with a model of the Epa9 adhesion domain comprising the full L1 region. This model was created with the Modeller interface of UCSF Chimera (Pettersen et al. 2004; Yang et al. 2012) using the lactose containing Epa9A structure (PDB 4CP0) as a template. Docking simulations were performed with AutoDock Vina to create an Epa9A model bound to the tetrameric galactoside gangliotetraose (Galβ1-3GalNAcβ1-4Galβ1-4Glc, PubChem CID: 90 475 902). All further steps for preparation and execution of molecular dynamics simulations were performed according to Lemkul 2019. In short, protein and ligand were first saved in two separate files and then parameterized using the CHARMM force field (Huang et al. 2017) implementation of GROMACS (Abraham et al. 2015) for the protein and the CHARMM General Force Field (CGENFF) tool (Vanommeslaeghe et al. 2010) for the ligand. Next, the generated topology files for both molecules were combined by including the file contents of the ligand topology file in the protein topology file. The combined protein-ligand structure was then placed in the center of a rhombic dodecahedral unit cell with a distance of 1 nm between the structure and the cell edges. A dodecahedric shape of the unit cell minimizes the amount of water necessary for solvation of the system as its volume is about 0.71 times that of a cubic cell unit. Thus choosing a rhombic dodecahedron reduces the computation time for the simulation. The protein-ligand complex was then solvated with water and additional sodium ions were added for a neutral net charge of the system. Afterwards, the system was relaxed by using an energy minimization step to create a reasonable starting structure. This was achieved by a short simulation run (max. $5 \cdot 10^4$ steps) to make sure that the system contains no steric clashes or inappropriate geometry. In the next step, the solvent and ions around the protein were equilibrated to a temperature of 300 K and a pressure of 10^5 Pa to ensure that the system would be stable during the simulation. After equilibration of temperature

and pressure, a molecular dynamics simulation of 10 ns was performed. The resulting trajectory was then processed to remove computational artifacts and recenter the protein in the unit cell. Afterwards, the simulated system was inspected visually and by calculation of the protein-ligand interaction energy. This energy comprises coulombic and van der Waals interactions between both molecules and thus provides a measure of possible nonbonded interactions between the Epa9A domain and the respective ligand. Additionally, hydrogen bonds between both molecules were calculated by using the *hbond* tool of GROMACS. The *hbond* tool calculates the distance between donor and acceptor and also considers the angle between donor, hydrogen and acceptor. This results in a list of all possible hydrogen bonds over the whole trajectory which is helpful in evaluating ligand binding and changes in protein conformation.

7 List of abbreviations

The units of measurement used in this work correspond to the International System of Units (SI; Système internationale d'unités) and units derived therefrom. Amino acids are identified by the one-letter code of the nomenclature of the "International Union of Biochemistry and Molecular Biology".

Å	Angstrom (1 Å = 100 pm = 10^{-10} m)
APS	Ammonium persulfate
BSA	Bovine serum albumin
C-terminal	Carboxy-terminal
CFG	Consortium for Functional Glycomics
CBL	Calcium binding loop
Da	Dalton (1 Da = 1 u = 1.660 539 040 · 10^{-27} kg)
DIC	Differential interference contrast
CWP	Cell wall-associated protein
DNA	Desoxyribonucleic acid
DNAseI	Desoxyribonuclease I
dNTP	Nucleoside triphosphate
EDTA	Ethylenediaminetetraacetic acid
Epa	Epithelial Adhesin
FTS	Fluorescence titration spectroscopy
Gal	D-galactose
Galα	Galactose linked via an α-glycosidic bond
Galα1-3Gal	α1-3-galactobiose
Galβ	Galactose linked via an β-glycosidic bond
Galβ1-3Gal	β1-3-galactobiose
Galβ1-3GalNAc	T-Antigen
Galβ1-3GlcNAc	Lacto-*N*-biose
Galβ1-4GlcNAc	*N*-acetyl-D-lactosamine
GalNAc	*N*-acetyl-D-galactosamine
Glc	D-glucose
GlcNAc	*N*-acetyl-D-glucosamine
GPI	Glycophosphatidylinositol
HDX-MS	Hydrogen–deuterium exchange mass spectrometry
kb	Kilobases
M	Molar (mol/l)
Man	Mannose
MDN	Molecular dynamics simulation
N-terminal	Amino-terminal

NTA	Nitrilotriacetic acid
NeuAc	*N*-acetylneuraminic acid
nt	Nucleotide
P	Promoter
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PGK1	Phosphoglycerate kinase 1
PMSF	Phenylmethylsulfonyl fluoride
rmsd	Root mean square deviation
rmsf	Root mean square fluctuation
rpm	Revolutions per minute
SC	Synthetic complete
T	Terminator
TDA	Terminal Disaccharide Analysis

8 List of Figures

9 List of Tables

10 References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **1-2**: 19–25. DOI: 10/f3pj2n.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). *PHENIX* : A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr D Biol Crystallogr* **66** (2): 213–221. DOI: 10/cftsm2.

Aouizerat, T., Gutman, I., Paz, Y., Maeir, A. M., Gadot, Y., Gelman, D., Szitenberg, A., Drori, E., Pinkus, A., Schoemann, M., Kaplan, R., Ben-Gedalya, T., Coppenhagen-Glazer, S., Reich, E., Saragovi, A., Lipschits, O., Klutstein, M., and Hazan, R. (2019). Isolation and Characterization of Live Yeast Cells from Ancient Vessels as a Tool in Bio-Archaeology. *mBio* **10** (2). DOI: 10/ggq599.

Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A., and Sebastian Seung, H. (2017). Trainable Weka Segmentation: A Machine Learning Tool for Microscopy Pixel Classification. *Bioinformatics* **33** (15): 2424–2426. DOI: 10/f9x7vt.

Balasubramanian, M. K., Bi, E., and Glotzer, M. (2004). Comparative Analysis of Cytokinesis in Budding Yeast, Fission Yeast and Animal Cells. *Curr. Biol.* **14** (18): R806–818. DOI: 10/dqrq7g.

Bartlett, A. H. and Park, P. W. (2010). Proteoglycans in Host-Pathogen Interactions: Molecular Mechanisms and Therapeutic Implications. *Expert Reviews in Molecular Medicine* **12**: e5. DOI: 10/br55gd.

Bauer, F. F., Govender, P., and Bester, M. C. (2010). Yeast Flocculation and Its Biotechnological Relevance. *Appl. Microbiol. Biotechnol.* **88** (1): 31–39. DOI: 10/fb39hq.

Bergstrom, K. S. B. and Xia, L. (2013). Mucin-Type O-Glycans and Their Roles in Intestinal Homeostasis. *Glycobiology* **23** (9): 1026–1037. DOI: 10/f45fb4.

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the Worldwide Protein Data Bank. *Nat. Struct. Biol.* **10** (12): 980. DOI: 10/fhdhcs.

Bester, M. C., Pretorius, I. S., and Bauer, F. F. (2006). The Regulation of *Saccharomyces cerevisiae FLO* Gene Expression and Ca^{2+} -Dependent Flocculation by Flo8p and Mss11p. *Curr. Genet.* **49** (6): 375–383. DOI: 10/b5zmvk.

Boisnard, S., Zhou Li, Y., Arnaise, S., Sequeira, G., Raffoux, X., Enache-Angoulvant, A., Bolotin-Fukuhara, M., and Fairhead, C. (2015). Efficient Mating-Type Switching in *Candida glabrata* Induces Cell Death. *PloS One* **10** (10): e0140990. DOI: 10/ghdx4k.

Briza, P., Breitenbach, M., Ellinger, A., and Segall, J. (1990). Isolation of Two Developmentally Regulated Genes Involved in Spore Wall Maturation in *Saccharomyces cerevisiae*. *Genes & Development* **4** (10): 1775–1789. DOI: 10/dwvc72.

Brockert, P. J., Lachke, S. A., Srikantha, T., Pujol, C., Galask, R., and Soll, D. R. (2003). Phenotypic Switching and Mating Type Switching of *Candida glabrata* at Sites of Colonization. *Infect. Immun.* **71** (12): 7109–7118. DOI: 10/dhz8zx.

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden Killers: Human Fungal Infections. *Sci Transl Med* **4** (165): 165rv13. DOI: 10/ghb47f.

Brückner, S. and Mösch, H.-U. (2012). Choosing the Right Lifestyle: Adhesion and Development in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* **36** (1): 25–58. DOI: 10/bjxb5b.

Brückner, S., Schubert, R., Kraushaar, T., Hartmann, R., Hoffmann, D., Jelli, E., Drescher, K., Müller, D. J., Essen, L.-O., and Mösch, H.-U. (2020). Kin Discrimination in Social Yeast Is Mediated by Cell Surface Receptors of the Flo11 Adhesin Family. *Elife* **9**. DOI: 10/ggr3zc.

Calvio, C., Romagnuolo, F., Vulcano, F., Speranza, G., and Morelli, C. F. (2018). Evidences on the Role of the Lid Loop of γ-Glutamyltransferases (GGT) in Substrate Selection. *Enzyme Microb. Technol.* **114**: 55–62. DOI: 10/gdmgc6.

Castanheira, M., Messer, S. A., Rhomberg, P. R., Dietrich, R. R., Jones, R. N., and Pfaller, M. A. (2014). Isavuconazole and Nine Comparator Antifungal Susceptibility Profiles for Common and Uncommon *Candida* Species Collected in 2012: Application of New CLSI Clinical Breakpoints and Epidemiological Cutoff Values. *Mycopathologia* **178** (1-2): 1–9. DOI: 10/ghcpq9.

Castaño, I., Pan, S.-J., Zupancic, M., Hennequin, C., Dujon, B., and Cormack, B. P. (2005). Telomere Length Control and Transcriptional Regulation of Subtelomeric Adhesins in *Candida glabrata*. *Mol. Microbiol.* **55** (4): 1246–1258. DOI: 10/cjgs2g.

Charlier, C., Nielsen, K., Daou, S., Brigitte, M., Chretien, F., and Dromer, F. (2009). Evidence of a Role for Monocytes in Dissemination and Brain Invasion by *Cryptococcus neoformans*. *Infect. Immun.* **77** (1): 120–127. DOI: 10/dcfxhr.

Coco, B. J., Bagg, J., Cross, L. J., Jose, A., Cross, J., and Ramage, G. (2008). Mixed *Candida albicans* and *Candida glabrata* Populations Associated with the Pathogenesis of Denture Stomatitis. *Oral Microbiology and Immunology* **23** (5): 377–383. DOI: 10/fks4k9.

Coronado, J. E., Mneimneh, S., Epstein, S. L., Qiu, W.-G., and Lipke, P. N. (2007). Conserved Processes and Lineage-Specific Proteins in Fungal Cell Wall Evolution. *Eukaryotic Cell* **6** (12): 2269–2277. DOI: 10/b8wjrs.

Csank, C. and Haynes, K. (2000). *Candida glabrata* Displays Pseudohyphal Growth. *FEMS Microbiol. Lett.* **189** (1): 115–120. DOI: 10/bg399w.

Cuéllar-Cruz, M., López-Romero, E., Villagómez-Castro, J. C., and Ruiz-Baca, E. (2012). *Candida* Species: New Insights into Biofilm Formation. *Future Microbiol* **7** (6): 755–771. DOI: 10/f32wg6.

De Groot, P. W. J., Ram, A. F., and Klis, F. M. (2005). Features and Functions of Covalently Linked Proteins in Fungal Cell Walls. *Fungal genetics and biology: FG & B* **42** (8): 657–675. DOI: 10/b4cw45.

De Las Peñas, A., Pan, S.-J., Castaño, I., Alder, J., Cregg, R., and Cormack, B. P. (2003). Virulence-Related Surface Glycoproteins in the Yeast Pathogen *Candida glabrata* Are Encoded in Subtelomeric Clusters and Subject to *RAP1*- and *SIR*-Dependent Transcriptional Silencing. *Genes Dev.* **17** (18): 2245–2258. DOI: 10/dt7t64.

De Marco, V., Stier, G., Blandin, S., and de Marco, A. (2004). The Solubility and Stability of Recombinant Proteins Are Increased by Their Fusion to NusA. *Biochem. Biophys. Res. Commun.* **322** (3): 766–771. DOI: 10/dvhh97.

De Groot, P. W. J., Ruiz, C., Vázquez de Aldana, C. R., Duenas, E., Cid, V. J., Del Rey, F., Rodríquez-Peña, J. M., Pérez, P., Andel, A., Caubín, J., Arroyo, J., García, J. C., Gil, C., Molina, M., García, L. J., Nombela, C., and Klis, F. M. (2001). A Genomic Approach for the Identification and Classification of Genes Involved in Cell Wall Formation and Its Regulation in *Saccharomyces cerevisiae*. *Comp. Funct. Genomics* **2** (3): 124–142. DOI: 10/cz4wv6.

De Groot, P. W. J., Bader, O., de Boer, A. D., Weig, M., and Chauhan, N. (2013). Adhesins in Human Fungal Pathogens: Glue with Plenty of Stick. *Eukaryotic Cell* **12** (4): 470–481. DOI: 10/gf5ch3.

De Groot, P. W. J., Kraneveld, E. A., Yin, Q. Y., Dekker, H. L., Gross, U., Crielaard, W., de Koster, C. G., Bader, O., Klis, F. M., and Weig, M. (2008). The Cell Wall of the Human Pathogen *Candida glabrata*: Differential Incorporation of Novel Adhesin-like Wall Proteins. *Eukaryotic Cell* **7** (11): 1951–1964. DOI: 10/d54tj4.

Del Caño-Ochoa, F., Grande-García, A., Reverte-López, M., D'Abramo, M., and Ramón-Maiques, S. (2018). Characterization of the Catalytic Flexible Loop in the Dihydroorotase Domain of the Human Multi-Enzymatic Protein CAD. *J. Biol. Chem.* **293** (49): 18903–18913. DOI: 10/gg3484.

De Melo Pereira, G. V., Soccol, V. T., Pandey, A., Medeiros, A. B. P., Andrade Lara, J. M. R., Gollo, A. L., and Soccol, C. R. (2014). Isolation, Selection and Evaluation of Yeasts for Use in Fermentation of Coffee Beans by the Wet Process. *Int. J. Food Microbiol.* **188**: 60–66. DOI: 10/f6rgdr.

Diderrich, R. (2014). "Strukturelle und funktionelle Charakterisierung der epithelialen Adhäsine aus *Candida glabrata*". Dissertation. Philipps-Universität Marburg. 243 pp.

Diderrich, R., Kock, M., Maestre-Reyna, M., Keller, P., Steuber, H., Rupp, S., Essen, L.-O., and Mösch, H.-U. (2015). Structural Hot Spots Determine Functional Diversity of the *Candida glabrata* Epithelial Adhesin Family. *J. Biol. Chem.* **290** (32): 19597–19613. DOI: 10/f7nvx4.

Diekema, D., Arbefeville, S., Boyken, L., Kroeger, J., and Pfaller, M. (2012). The Changing Epidemiology of Healthcare-Associated Candidemia over Three Decades. *Diagn. Microbiol. Infect. Dis.* **73** (1): 45–48. DOI: 10/f3z8d5.

Dieterich, C., Schandar, M., Noll, M., Johannes, F.-J., Brunner, H., Graeve, T., and Rupp, S. (2002). *In Vitro* Reconstructed Human Epithelia Reveal Contributions of *Candida albicans EFG1* and *CPH1* to Adhesion and Invasion. *Microbiology (Reading, Engl.)* **148** (Pt 2): 497–506. DOI: 10/gg3gvf.

Douglas, L. J. (2003). *Candida* Biofilms and Their Role in Infection. *Trends Microbiol.* **11** (1): 30–36. DOI: 10/c3ngr5.

Drenth, J. and Mesters, J. (2007). *Principles of Protein X-Ray Crystallography*. 3. ed. Literaturverz. S. 316 - 325. - Index. New York, NY: Springer. 332 pp.

Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., Ji, X.-L., and Liu, S.-Q. (2016). Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int J Mol Sci* **17** (2). DOI: 10.3390/ijms17020144.

Dupres, V., Alsteens, D., Wilk, S., Hansen, B., Heinisch, J. J., and Dufrêne, Y. F. (2009). The Yeast Wsc1 Cell Surface Sensor Behaves like a Nanospring *in Vivo*. *Nature Chemical Biology* **5** (11): 857–862. DOI: 10/bcdkdx.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and Development of *Coot*. *Acta Crystallogr D Biol Crystallogr* **66** (4): 486–501. DOI: 10/bpv352.

Essen, L.-O., Vogt, M. S., and Mösch, H.-U. (2020). Diversity of GPI-Anchored Fungal Adhesins. *Biol Chem*. DOI: 10/ghgh3w.

Evans, P. R. (2011). An Introduction to Data Reduction: Space-Group Determination, Scaling and Intensity Statistics. *Acta Crystallogr D Biol Crystallogr* **67** (4): 282–292. DOI: 10/fk86qj.

Fabre, E., Muller, H., Therizols, P., Lafontaine, I., Dujon, B., and Fairhead, C. (2005). Comparative Genomics in Hemiascomycete Yeasts: Evolution of Sex, Silencing, and Subtelomeres. *Mol. Biol. Evol.* **22** (4): 856–873. DOI: 10/fh26v4.

Fichtner, L., Schulze, F., and Braus, G. H. (2007). Differential Flo8p-Dependent Regulation of *FLO1* and *FLO11* for Cell-Cell and Cell-Substrate Adherence of *S. cerevisiae* S288c. *Mol. Microbiol.* **66** (5): 1276–1289. DOI: 10/d7cm6q.

Fidel, P. L., Vazquez, J. A., and Sobel, J. D. (1999). *Candida glabrata*: Review of Epidemiology, Pathogenesis, and Clinical Disease with Comparison to *C. albicans*. *Clin. Microbiol. Rev.* **12** (1): 80–96. DOI: 10/ghdx76.

Flores, C.-L., Rodríguez, C., Petit, T., and Gancedo, C. (2000). Carbohydrate and Energy-Yielding Metabolism in Non-Conventional Yeasts. *FEMS Microbiol Rev* **24** (4): 507–529. DOI: 10/fc2kc4.

Free, S. J. (2013). Fungal Cell Wall Organization and Biosynthesis. *Adv. Genet.* **81**: 33–82. DOI: 10/gf5cj2.

Friederichs, S. (2018). "Der Einfluss des L1- und CBL2-Loops der Epa-Proteine aus der humanpathogenen Hefe *Candida glabrata* auf die Ligandenbindespezifität". Master thesis. Philipps-Universität Marburg. 76 pp.

Frieman, M. B., McCaffery, J. M., and Cormack, B. P. (2002). Modular Domain Structure in the *Candida glabrata Adhesin* Epa1p, a Beta1,6 Glucan-Cross-Linked Cell Wall Protein. *Mol. Microbiol.* **46** (2): 479–492. DOI: 10/fgvgr4.

Gabaldón, T. and Carreté, L. (2016). The Birth of a Deadly Yeast: Tracing the Evolutionary Emergence of Virulence Traits in *Candida glabrata*. *FEMS Yeast Res.* **16** (2): fov110. DOI: 10/f8f9ht.

Gabaldón, T. and Fairhead, C. (2019). Genomes Shed Light on the Secret Life of *Candida glabrata*: Not so Asexual, Not so Commensal. *Curr. Genet.* **65** (1): 93–98. DOI: 10/gf5ckf.

Gabaldón, T., Martin, T., Marcet-Houben, M., Durrens, P., Bolotin-Fukuhara, M., Lespinet, O., Arnaise, S., Boisnard, S., Aguileta, G., Atanasova, R., Bouchier, C., Couloux, A., Creno, S., Almeida Cruz, J., Devillers, H., Enache-Angoulvant, A., Guitard, J., Jaouen, L., Ma, L., Marck, C., Neuvéglise, C., Pelletier, E., Pinard, A., Poulain, J., Recoquillay, J., Westhof, E., Wincker, P., Dujon, B., Hennequin, C., and Fairhead, C. (2013). Comparative Genomics of Emerging Pathogens in the *Candida glabrata* Clade. *BMC Genomics* **14**: 623. DOI: 10/gb3gjt.

Ganguly, S. and Mitchell, A. P. (2011). Mucosal Biofilms of *Candida albicans*. *Curr. Opin. Microbiol.* **14** (4): 380–385. DOI: 10/c6n6t2.

García, R., Bermejo, C., Grau, C., Pérez, R., Rodríguez-Peña, J. M., Francois, J., Nombela, C., and Arroyo, J. (2004). The Global Transcriptional Response to Transient Cell Wall Damage in *Saccharomyces cerevisiae* and Its Regulation by the Cell Integrity Signaling Pathway. *The Journal of Biological Chemistry* **279** (15): 15183–15195. DOI: 10/fchfgz.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). "Protein Identification and Analysis Tools on the ExPASy Server". In: *The Proteomics Protocols Handbook*. Ed. by J. M. Walker. Totowa, NJ: Humana Press: 571–607. DOI: 10.1385/1-59259-890-0:571.

Geissner, A., Reinhardt, A., Rademacher, C., Johannssen, T., Monteiro, J., Lepenies, B., Thépaut, M., Fieschi, F., Mrázková, J., Wimmerova, M., Schuhmacher, F., Götze, S., Grünstein, D., Guo, X., Hahm, H. S., Kandasamy, J., Leonori, D., Martin, C. E., Parameswarappa, S. G., Pasari, S., Schlegel, M. K., Tanaka, H., Xiao, G., Yang, Y., Pereira, C. L., Anish, C., and Seeberger, P. H. (2019). Microbe-Focused Glycan Array Screening Platform. *Proc. Natl. Acad. Sci. U.S.A.* **116** (6): 1958–1967. DOI: 10/ghbfj4.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A. P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.-D., Flaherty, P., Foury, F., Garfinkel, D. J., Gerstein, M., Gotte, D., Güldener, U., Hegemann, J. H., Hempel, S., Herman, Z., Jaramillo, D. F., Kelly, D. E., Kelly, S. L., Kötter, P., LaBonte, D., Lamb, D. C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S. L., Revuelta, J. L., Roberts, C. J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D. D., Sookhai-Mahadeo, S., Storms, R. K., Strathern, J. N., Valle, G., Voet, M., Volckaert, G., Wang, C.-y., Ward, T. R., Wilhelmy, J., Winzeler, E. A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J. D., Snyder, M., Philippsen, P., Davis, R. W., and Johnston, M. (2002). Functional Profiling of the *Saccharomyces cerevisiae* Genome. *Nature* **418** (6896): 387–391. DOI: 10/dhbs45.

Gil, N. F., Martinez, R. C. R., Gomes, B. C., Nomizo, A., and De Martinis, E. C. P. (2010). Vaginal Lactobacilli as Potential Probiotics against *Candida* spp. *Braz. J. Microbiol.* **41** (1): 6–14. DOI: 10/bhvws9.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* **19** (1): 163–164. DOI: 10/d795gr.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996). Life with 6000 Genes. *Science* **274** (5287): 546, 563–567. DOI: 10/d7mhxv.

Gow, N. A. R., Latge, J.-P., and Munro, C. A. (2017). The Fungal Cell Wall: Structure, Biosynthesis, and Function. *Microbiol Spectr* **5** (3). DOI: 10/gf88gw.

Green, J. V., Orsborn, K. I., Zhang, M., Tan, Q. K. G., Greis, K. D., Porollo, A., Andes, D. R., Long Lu, J., and Hostetter, M. K. (2013). Heparin-Binding Motifs and Biofilm Formation by *Candida albicans*. *The Journal of Infectious Diseases* **208** (10): 1695–1704. DOI: 10/f5f5jr.

Grossart, H.-P., Van den Wyngaert, S., Kagami, M., Wurzbacher, C., Cunliffe, M., and Rojas-Jimenez, K. (2019). Fungi in Aquatic Ecosystems. *Nat. Rev. Microbiol.* **17** (6): 339–354. DOI: 10/gf2kvb.

Guinea, J. (2014). Global Trends in the Distribution of *Candida* Species Causing Candidemia. *Clin. Microbiol. Infect.* **20 Suppl 6**: 5–10. DOI: 10/f5699c.

Hall, R. A. and Noverr, M. C. (2017). Fungal Interactions with the Human Host: Exploring the Spectrum of Symbiosis. *Curr. Opin. Microbiol.* **40**: 58–64. DOI: 10/gcrfnv.

Heimburg-Molinaro, J., Song, X., Smith, D. F., and Cummings, R. D. (2011). Preparation and Analysis of Glycan Microarrays. *Curr Protoc Protein Sci* **Chapter 12**: Unit12.10. DOI: 10/csn423.

Henderson, R. (1990). Cryo-Protection of Protein Crystals against Radiation Damage in Electron and X-Ray Diffraction. *Proc. R. Soc. Lond. B* **241** (1300): 6–8. DOI: 10/cqq8fn.

Hickey, W. F., Sommerville, L. H., and Schoen, F. J. (1983). Disseminated *Candida glabrata*: Report of a Uniquely Severe Infection and a Literature Review. *Am. J. Clin. Pathol.* **80** (5): 724–727. DOI: 10/ghcp2r.

Hoffmann, D., Diderrich, R., Reithofer, V., Friederichs, S., Kock, M., Essen, L.-O., and Mösch, H.-U. (2020). Functional Reprogramming of *Candida glabrata* Epithelial Adhesins: The Role of Conserved and Variable Structural Motifs in Ligand Binding. *J. Biol. Chem.* **295** (35): 12512–12524. DOI: 10/gg6r5w.

Hu, D., Tateno, H., and Hirabayashi, J. (2015). Lectin Engineering, a Molecular Evolutionary Approach to Expanding the Lectin Utilities. *Molecules* **20** (5): 7637–7656. DOI: 10/f7hp6n.

Hu, D., Tateno, H., Kuno, A., Yabe, R., and Hirabayashi, J. (2012). Directed Evolution of Lectins with Sugar-Binding Specificity for 6-Sulfo-Galactose. *J. Biol. Chem.* **287** (24): 20313–20320. DOI: 10/f32g7t.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., and MacKerell, A. D. (2017). CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **14** (1): 71–73. DOI: 10/gfxkj3.

Ielasi, F. S., Decanniere, K., and Willaert, R. G. (2012). The Epithelial Adhesin 1 (Epa1p) from the Human-Pathogenic Yeast *Candida glabrata*: Structural and Functional Study of the Carbohydrate-Binding Domain. *Acta Crystallogr. D Biol. Crystallogr.* **68** (Pt 3): 210–217. DOI: 10/gf5ch8.

Ielasi, F. S., Verhaeghe, T., Desmet, T., and Willaert, R. G. (2014). Engineering the Carbohydrate-Binding Site of Epa1p from *Candida glabrata*: Generation of Adhesin Mutants with Different Carbohydrate Specificity. *Glycobiology* **24** (12): 1312–1322. DOI: 10/f6vxsk.

Jancarik, J. and Kim, S. H. (1991). Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins. *J Appl Crystallogr* **24** (4): 409–411. DOI: 10/c2nhm2.

Jentoft, N. (1990). Why Are Proteins O-Glycosylated? *Trends Biochem. Sci.* **15** (8): 291–294. DOI: 10/c64w7k.

Junker, K., Bravo Ruiz, G., Lorenz, A., Walker, L., Gow, N. A. R., and Wendland, J. (2018). The Mycoparasitic Yeast *Saccharomycopsis schoenii* Predates and Kills Multi-Drug Resistant *Candida auris*. *Sci Rep* **8** (1): 14959. DOI: 10/ghbts8.

Kabsch, W. (2010a). Integration, Scaling, Space-Group Assignment and Post-Refinement. *Acta Crystallogr D Biol Crystallogr* **66** (2): 133–144. DOI: 10/bhb84x.

Kabsch, W. (2010b). *XDS*. *Acta Crystallogr D Biol Crystallogr* **66** (2): 125–132. DOI: 10/csc47s.

Kasper, L., Seider, K., and Hube, B. (2015). Intracellular Survival of *Candida glabrata* in Macrophages: Immune Evasion and Persistence. *FEMS Yeast Res.* **15** (5): fov042. DOI: 10/f7sgdn.

Kaur, R., Domergue, R., Zupancic, M. L., and Cormack, B. P. (2005). A Yeast by Any Other Name: *Candida glabrata* and Its Interaction with the Host. *Curr. Opin. Microbiol.* **8** (4): 378–384. DOI: 10/fkfnd8.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis. *Nature* **181** (4610): 662–666. DOI: 10/drvnxh.

Kleiman, M. and Tannenbaum, E. (2009). Diploidy and the Selective Advantage for Sexual Reproduction in Unicellular Organisms. *Theory Biosci.* **128** (4): 249–285. DOI: 10/fqw23s.

Klis, F. M., de Groot, P., and Hellingwerf, K. (2001). Molecular Organization of the Cell Wall of *Candida albicans*. *Med. Mycol.* **39 Suppl 1**: 1–8. DOI: 10/djxtdg.

Klotz, S. A., Chasin, B. S., Powell, B., Gaur, N. K., and Lipke, P. N. (2007). Polymicrobial Bloodstream Infections Involving *Candida* Species: Analysis of Patients and Review of the Literature. *Diagn. Microbiol. Infect. Dis.* **59** (4): 401–406. DOI: 10/fnxhbw.

Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K., and Schiebel, E. (1999). Epitope Tagging of Yeast Genes Using a PCR-Based Strategy: More Tags and Improved Practical Routines. *Yeast* **15** (10B): 963–972. DOI: 10/fkx4c8.

Kock, C., Dufrêne, Y. F., and Heinisch, J. J. (2015). Up against the Wall: Is Yeast Cell Wall Integrity Ensured by Mechanosensing in Plasma Membrane Microdomains? *Appl. Environ. Microbiol.* **81** (3): 806–811. DOI: 10/f6wfbg.

Kock, M. A. (2015). "Vergleichende Struktur-/Funktionsanalyse von Adhäsionsdomänen pathogener und nicht-pathogener Hefen". Dissertation. Philipps-Universität Marburg. 194 pp.

Kojic, E. M. and Darouiche, R. O. (2004). *Candida* Infections of Medical Devices. *Clin. Microbiol. Rev.* **17** (2): 255–267. DOI: 10/bz5jw4.

Kollár, R., Reinhold, B. B., Petráková, E., Yeh, H. J., Ashwell, G., Drgonová, J., Kapteyn, J. C., Klis, F. M., and Cabib, E. (1997). Architecture of the Yeast Cell Wall. Beta(1–>6)-Glucan Interconnects Mannoprotein, Beta(1–>)3-Glucan, and Chitin. *J. Biol. Chem.* **272** (28): 17762–17775. DOI: 10/dwxs77.

Kraneveld, E. A., de Soet, J. J., Deng, D. M., Dekker, H. L., de Koster, C. G., Klis, F. M., Crielaard, W., and de Groot, P. W. J. (2011). Identification and Differential Gene Expression of Adhesin-like Wall Proteins in *Candida glabrata* Biofilms. *Mycopathologia* **172** (6): 415–427. DOI: 10/d7w8vc.

Kraushaar, T. (2016). "Strukturelle und biophysikalische Analyse von Adhäsionsdomänen des Flo11-Typs aus Ascomyceten". Dissertation. Philipps-Universität Marburg. 213 pp.

Kraushaar, T., Brückner, S., Veelders, M., Rhinow, D., Schreiner, F., Birke, R., Pagenstecher, A., Mösch, H.-U., and Essen, L.-O. (2015). Interactions by the Fungal Flo11 Adhesin Depend on

a Fibronectin Type III-like Adhesin Domain Girdled by Aromatic Bands. *Structure* **23** (6): 1005–1017. DOI: 10/f7d26r.

Kühlbrandt, W. (2014). Cryo-EM Enters a New Era. *Elife* **3**: e03678. DOI: 10/ggfkr6.

Kuhn, D. M. and Vyas, V. K. (2012). The *Candida glabrata* Adhesin Epa1p Causes Adhesion, Phagocytosis, and Cytokine Secretion by Innate Immune Cells. *FEMS Yeast Res.* **12** (4): 398–414. DOI: 10/fxprn6.

Kurtzman, C. P. and Piškur, J. (2006). "Taxonomy and Phylogenetic Diversity among the Yeasts". In: *Comparative Genomics*. Ed. by P. Sunnerhagen and J. Piskur. Vol. 15. Topics in Current Genetics. Berlin, Heidelberg: Springer Berlin Heidelberg: 29–46. DOI: 10.1007/b106654.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227** (5259): 680–685. DOI: 10/br72fg.

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005). ConSurf 2005: The Projection of Evolutionary Conservation Scores of Residues on Protein Structures. *Nucleic Acids Res.* **33** (Web Server issue): W299–302. DOI: 10/b7r3dw.

Lee, D. H., Oh, J.-H., and Chung, J. H. (2016). Glycosaminoglycan and Proteoglycan in Skin Aging. *Journal of Dermatological Science* **83** (3): 174–181. DOI: 10/f83s9h.

Lemkul, J. (2019). From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. *LiveCoMS* **1** (1). DOI: 10/ggwjxw.

Lesage, G. and Bussey, H. (2006). Cell Wall Assembly in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **70** (2): 317–343. DOI: 10/cbdfc7.

Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., and Tan, S. (2005). Comparison of Affinity Tags for Protein Purification. *Protein Expr. Purif.* **41** (1): 98–105. DOI: 10/dqkmkm.

Lin, J., Oh, S.-H., Jones, R., Garnett, J. A., Salgado, P. S., Rusnakova, S., Matthews, S. J., Hoyer, L. L., and Cota, E. (2014). The Peptide-Binding Cavity Is Essential for Als3-Mediated Adhesion of *Candida albicans* to Human Cells. *J. Biol. Chem.* **289** (26): 18401–18412. DOI: 10/ggm84h.

Lipfert, J. and Doniach, S. (2007). Small-Angle X-Ray Scattering from RNA, Proteins, and Protein Complexes. *Annu Rev Biophys Biomol Struct* **36**: 307–327. DOI: 10/fg2b9t.

Lipke, P. N. and Ovalle, R. (1998). Cell Wall Architecture in Yeast: New Structure and New Challenges. *J. Bacteriol.* **180** (15): 3735–3740. DOI: 10/ghcz85.

Liti, G. (2015). The Fascinating and Secret Wild Life of the Budding Yeast *S. cerevisiae*. *Elife* **4**. DOI: 10/gf5knc.

Liu, G., Shen, Y., Atreya, H. S., Parish, D., Shao, Y., Sukumaran, D. K., Xiao, R., Yee, A., Lemak, A., Bhattacharya, A., Acton, T. A., Arrowsmith, C. H., Montelione, G. T., and Szyperski, T. (2005). NMR Data Collection and Analysis Protocol for High-Throughput Protein Structure Determination. *Proc. Natl. Acad. Sci. U.S.A.* **102** (30): 10487–10492. DOI: 10/c4s5vb.

Ludewig, L. (2013). "Funktionelle und strukturelle Charakterisierung von Epithelzell-Adhäsinen aus *C. glabrata*". Master thesis. Philipps-Universität Marburg. 64 pp.

Lutterbach, B. J. (2019). "Structural and Functional Characterization of Fungal Cell Wall Proteins Involved in Adhesion and Integrity Sensing". Dissertation. Philipps-Universität Marburg. 173 pp.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019). The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* **47** (W1): W636–W641. DOI: 10/gf73xm.

Madzak, C. (2018). Engineering *Yarrowia lipolytica* for Use in Biotechnological Applications: A Review of Major Achievements and Recent Innovations. *Mol. Biotechnol.* **60** (8): 621–635. DOI: 10/ghbzhx.

Maestre-Reyna, M., Diderrich, R., Veelders, M. S., Eulenburg, G., Kalugin, V., Brückner, S., Keller, P., Rupp, S., Mösch, H.-U., and Essen, L.-O. (2012). Structural Basis for Promiscuity and Specificity during *Candida glabrata* Invasion of Host Epithelia. *Proc. Natl. Acad. Sci. U.S.A.* **109** (42): 16864–16869. DOI: 10/f4c89m.

Massey, S. E., Moura, G., Beltrão, P., Almeida, R., Garey, J. R., Tuite, M. F., and Santos, M. A. S. (2003). Comparative Evolutionary Genomics Unveils the Molecular Mechanism of Reassignment of the CTG Codon in *Candida* Spp. *Genome Res.* **13** (4): 544–557. DOI: 10/fgg66f.

McCoy, A. J. (2007). Solving Structures of Protein Complexes by Molecular Replacement with *Phaser*. *Acta Crystallogr D Biol Crystallogr* **63** (1): 32–41. DOI: 10/fd9cpx.

Merico, A., Sulo, P., Piskur, J., and Compagno, C. (2007). Fermentative Lifestyle in Yeasts Belonging to the *Saccharomyces* Complex. *FEBS J.* **274** (4): 976–989. DOI: 10/cbw7d7.

Mills, J. E. and Dean, P. M. (1996). Three-Dimensional Hydrogen-Bond Geometry and Probability Information from a Crystal Survey. *J. Comput. Aided Mol. Des.* **10** (6): 607–622. DOI: 10/cqtvxm.

Morla, S. (2019). Glycosaminoglycans and Glycosaminoglycan Mimetics in Cancer and Inflammation. *International Journal of Molecular Sciences* **20** (8). DOI: 10/ghdwss.

Negri, M., Botelho, C., Silva, S., Lopes, L. M. R. H., Henriques, M., Azeredo, J., and Oliveira, R. (2011). An *in Vitro* Evaluation of *Candida tropicalis* Infectivity Using Human Cell Monolayers. *J. Med. Microbiol.* **60** (Pt 9): 1270–1275. DOI: 10/dhxks6.

Nguyen, N. H., Suh, S.-O., and Blackwell, M. (2007). Five Novel *Candida* Species in Insect-Associated Yeast Clades Isolated from Neuroptera and Other Insects. *Mycologia* **99** (6): 842–858. DOI: 10/bxmrkb.

Nieuw Amerongen, A. V., Bolscher, J. G., Bloemena, E., and Veerman, E. C. (1998). Sulfomucins in the Human Body. *Biol. Chem.* **379** (1): 1–18. DOI: 10/bfhtgh.

Paulick, M. G. and Bertozzi, C. R. (2008). The Glycosylphosphatidylinositol Anchor: A Complex Membrane-Anchoring Structure for Proteins. *Biochemistry* **47** (27): 6991–7000. DOI: 10/bxbzs6.

Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H., and Liddington, R. C. (1997). Crystal Structure of the Anthrax Toxin Protective Antigen. *Nature* **385** (6619): 833–838. DOI: 10/d3p4f6.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera–a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **25** (13): 1605–1612. DOI: 10/b4bq4c.

Pfaller, M. A., Diekema, D. J., Turnidge, J. D., Castanheira, M., and Jones, R. N. (2019). Twenty Years of the SENTRY Antifungal Surveillance Program: Results for *Candida* Species From 1997-2016. *Open Forum Infect Dis* **6** (Suppl 1): S79–S94. DOI: 10/ggh925.

Pfaller, M. A., Messer, S. A., Moet, G. J., Jones, R. N., and Castanheira, M. (2011). *Candida* Bloodstream Infections: Comparison of Species Distribution and Resistance to Echinocandin and Azole Antifungal Agents in Intensive Care Unit (ICU) and Non-ICU Settings in the SENTRY Antimicrobial Surveillance Program (2008-2009). *Int. J. Antimicrob. Agents* **38** (1): 65–69. DOI: 10/dpnw5r.

Redding, S. W., Dahiya, M. C., Kirkpatrick, W. R., Coco, B. J., Patterson, T. F., Fothergill, A. W., Rinaldi, M. G., and Thomas, C. R. (2004). *Candida glabrata* Is an Emerging Cause of Oropharyngeal Candidiasis in Patients Receiving Radiation for Head and Neck Cancer. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* **97** (1): 47–52. DOI: 10/ds5zxq.

Redding, S. W., Kirkpatrick, W. R., Dib, O., Fothergill, A. W., Rinaldi, M. G., and Patterson, T. F. (2000). The Epidemiology of Non-*albicans Candida* in Oropharyngeal Candidiasis in HIV Patients. *Special Care in Dentistry* **20** (5): 178–181. DOI: 10/c9n8x7.

Rigden, D. J., Mello, L. V., and Galperin, M. Y. (2004). The PA14 Domain, a Conserved All-Beta Domain in Bacterial Toxins, Enzymes, Adhesins and Signaling Molecules. *Trends Biochem. Sci.* **29** (7): 335–339. DOI: 10/dg3k56.

Rodrigues, C. F., Silva, S., and Henriques, M. (2014). *Candida glabrata*: A Review of Its Features and Resistance. *Eur. J. Clin. Microbiol. Infect. Dis.* **33** (5): 673–688. DOI: 10/f5zvj6.

Roetzer, A., Gregori, C., Jennings, A. M., Quintin, J., Ferrandon, D., Butler, G., Kuchler, K., Ammerer, G., and Schüller, C. (2008). *Candida glabrata* Environmental Stress Response Involves *Saccharomyces cerevisiae* Msn2/4 Orthologous Transcription Factors. *Mol. Microbiol.* **69** (3): 603–620. DOI: 10/dngvf3.

Rostand, K. S. and Esko, J. D. (1997). Microbial Adherence to and Invasion through Proteoglycans. *Infection and Immunity* **65** (1): 1–8. DOI: 10/ghdwz5.

Salvato-Vallerdu, G. (2009). *The Perrin - Jablonski Diagram*. texsample.net. URL: http://www.texample.net/tikz/examples/the-perrin-jablonski-diagram/ (visited on 2020-03-10).

Sherman, F. (1991). Getting Started with Yeast. *Meth. Enzymol.* **194**: 3–21. DOI: 10/fr7mdr.

Sherman, F., Fink, G. R., and Hicks, J. B. (1987). *Laboratory Course Manual for Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press.

Sims, C. R., Ostrosky-Zeichner, L., and Rex, J. H. (2005). Invasive Candidiasis in Immunocompromised Hospitalized Patients. *Arch. Med. Res.* **36** (6): 660–671. DOI: 10/bz8qx3.

Smith, D. F. and Cummings, R. D. (2013). Application of Microarrays for Deciphering the Structure and Function of the Human Glycome. *Mol. Cell Proteomics* **12** (4): 902–912. DOI: 10/gg894z.

Soares, E. V. (2011). Flocculation in *Saccharomyces cerevisiae*: A Review. *J. Appl. Microbiol.* **110** (1): 1–18. DOI: 10/dk39gv.

Stank, A., Kokh, D. B., Fuller, J. C., and Wade, R. C. (2016). Protein Binding Pocket Dynamics. *Acc. Chem. Res.* **49** (5): 809–815. DOI: 10/f8mr6x.

Stevens, R. C., Yokoyama, S., and Wilson, I. A. (2001). Global Efforts in Structural Genomics. *Science* **294** (5540): 89–92. DOI: 10/fm54nx.

Tam, P., Gee, K., Piechocinski, M., and Macreadie, I. (2015). *Candida glabrata*, Friend and Foe. *J Fungi (Basel)* **1** (2): 277–292. DOI: 10/gf5chz.

Tanaka-Okamoto, M., Mukai, M., Takahashi, H., Fujiwara, Y., Ohue, M., and Miyamoto, Y. (2017). Various Sulfated Carbohydrate Tumor Marker Candidates Identified by Focused Glycomic Analyses. *Glycobiology* **27** (5): 400–415. DOI: 10/f93mb5.

Tati, S., Davidow, P., McCall, A., Hwang-Wong, E., Rojas, I. G., Cormack, B., and Edgerton, M. (2016). *Candida glabrata* Binding to *Candida albicans* Hyphae Enables Its Development in Oropharyngeal Candidiasis. *PLoS Pathog.* **12** (3): e1005522. DOI: 10/f8sh5k.

Timmermans, B., De Las Peñas, A., Castaño, I., and Van Dijck, P. (2018). Adhesins in *Candida glabrata*. *J Fungi (Basel)* **4** (2). DOI: 10/gf5b57.

Trott, O. and Olson, A. J. (2010). AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **31** (2): 455–461. DOI: 10/fqnzrb.

Uwamahoro, N., Verma-Gaur, J., Shen, H.-H., Qu, Y., Lewis, R., Lu, J., Bambery, K., Masters, S. L., Vince, J. E., Naderer, T., and Traven, A. (2014). The Pathogen *Candida albicans* Hijacks Pyroptosis for Escape from Macrophages. *mBio* **5** (2): e00003–00014. DOI: 10/ggj6mj.

Valotteau, C., Prystopiuk, V., Cormack, B. P., and Dufrêne, Y. F. (2019). Atomic Force Microscopy Demonstrates That *Candida glabrata* Uses Three Epa Proteins To Mediate Adhesion to Abiotic Surfaces. *mSphere* **4** (3). DOI: 10/gf5cj9.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell, A. D. (2010). CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comput Chem* **31** (4): 671–690. DOI: 10/fgkjs4.

Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F., Lütteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K. F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015). Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25** (12): 1323–1324. DOI: 10/ggpk8v.

Vazquez, J. A. (1999). Options for the Management of Mucosal Candidiasis in Patients with AIDS and HIV Infection. *Pharmacotherapy* **19** (1): 76–87. DOI: 10/frq5bt.

Veelders, M., Brückner, S., Ott, D., Unverzagt, C., Mösch, H.-U., and Essen, L.-O. (2010). Structural Basis of Flocculin-Mediated Social Behavior in Yeast. *Proc. Natl. Acad. Sci. U.S.A.* **107** (52): 22511–22516. DOI: 10/bp6j75.

Vogt, M. S., Schmitz, G. F., Varón Silva, D., Mösch, H.-U., and Essen, L.-O. (2020). Structural Base for the Transfer of GPI-Anchored Glycoproteins into Fungal Cell Walls. *Proc. Natl. Acad. Sci. U.S.A.* **117** (36): 22061–22067. DOI: 10/ghc9ps.

Weiner, M. P., Costa, G. L., Schoettlin, W., Cline, J., Mathur, E., and Bauer, J. C. (1994). Site-Directed Mutagenesis of Double-Stranded DNA by the Polymerase Chain Reaction. *Gene* **151** (1-2): 119–123. DOI: 10/cgdm9g.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011). Overview of the *CCP* 4 Suite and Current Developments. *Acta Crystallogr D Biol Crystallogr* **67** (4): 235–242. DOI: 10/dv3zq9.

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., Chu, A. M., Connelly, C., Davis, K., Dietrich, F., Dow, S. W., El Bakkoury, M., Foury, F., Friend, S. H., Gentalen, E., Giaever, G., Hegemann, J. H., Jones, T., Laub, M., Liao, H., Liebundguth, N., Lockhart, D. J., Lucau-Danila, A., Lussier, M., M'Rabet, N., Menard, P., Mittmann, M., Pai, C., Rebischung, C., Revuelta, J. L., Riles, L., Roberts, C. J., Ross-MacDonald, P., Scherens, B., Snyder, M., Sookhai-Mahadeo, S., Storms, R. K., Véronneau, S., Voet, M., Volckaert, G., Ward, T. R., Wysocki, R., Yen, G. S., Yu, K., Zimmermann, K., Philippsen, P., Johnston, M., and Davis, R. W. (1999). Functional Characterization of the *S. cerevisiae* Genome by Gene Deletion and Parallel Analysis. *Science* **285** (5429): 901–906. DOI: 10/cmrrvc.

Wüthrich, K. (1990). Protein Structure Determination in Solution by NMR Spectroscopy. *J. Biol. Chem.* **265** (36): 22059–22062.

Wüthrich, K. (2001). The Way to NMR Structures of Proteins. *Nat. Struct. Biol.* **8** (11): 923–925. DOI: 10/b2mwnc.

Xiao, X. and Lowe, M. E. (2015). The *B*5-Loop and Lid Domain Contribute to the Substrate Specificity of Pancreatic Lipase-Related Protein 2 (PNLIPRP2). *J. Biol. Chem.* **290** (48): 28847–28856. DOI: 10/gg348p.

Xu, Z., Green, B., Benoit, N., Schatz, M., Wheelan, S., and Cormack, B. (2020). *De Novo* Genome Assembly of *Candida glabrata* Reveals Cell Wall Protein Complement and Structure of Dispersed Tandem Repeat Arrays. *Mol. Microbiol.* DOI: 10/ggqtj9.

Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C. C., Pettersen, E. F., Goddard, T. D., Meng, E. C., Sali, A., and Ferrin, T. E. (2012). UCSF Chimera, MODELLER, and IMP: An Integrated Modeling System. *J. Struct. Biol.* **179** (3): 269–278. DOI: 10/fhq7pt.

Al-Yasiri, M. H., Normand, A.-C., L'Ollivier, C., Lachaud, L., Bourgeois, N., Rebaudet, S., Piarroux, R., Mauffrey, J.-F., and Ranque, S. (2016). Opportunistic Fungal Pathogen *Candida glabrata* Circulates between Humans and Yellow-Legged Gulls. *Sci Rep* **6**: 36157. DOI: 10/f885c3.

Zajac, D., Karkowska-Kuleta, J., Bochenska, O., Rapala-Kozik, M., and Kozik, A. (2016). Interaction of Human Fibronectin with *Candida glabrata* Epithelial Adhesin 6 (Epa6). *Acta Biochim. Pol.* **63** (3): 417–426. DOI: 10/f8959q.

Zhou, X., Haudenschild, A. K., Sherlock, B. E., Hu, J. C., Leach, J. K., Athanasiou, K. A., and Marcu, L. (2018). Detection of Glycosaminoglycan Loss in Articular Cartilage by Fluorescence Lifetime Imaging. *Journal of Biomedical Optics* **23** (12): 1–8. DOI: 10/ghdwmf.

Zupancic, M. L., Frieman, M., Smith, D., Alvarez, R. A., Cummings, R. D., and Cormack, B. P. (2008). Glycan Microarray Analysis of *Candida glabrata* Adhesin Ligand Specificity. *Mol. Microbiol.* **68** (3): 547–559. DOI: 10/bt3xrd.

11 Appendix

11.1 Fluorescence titration measurements

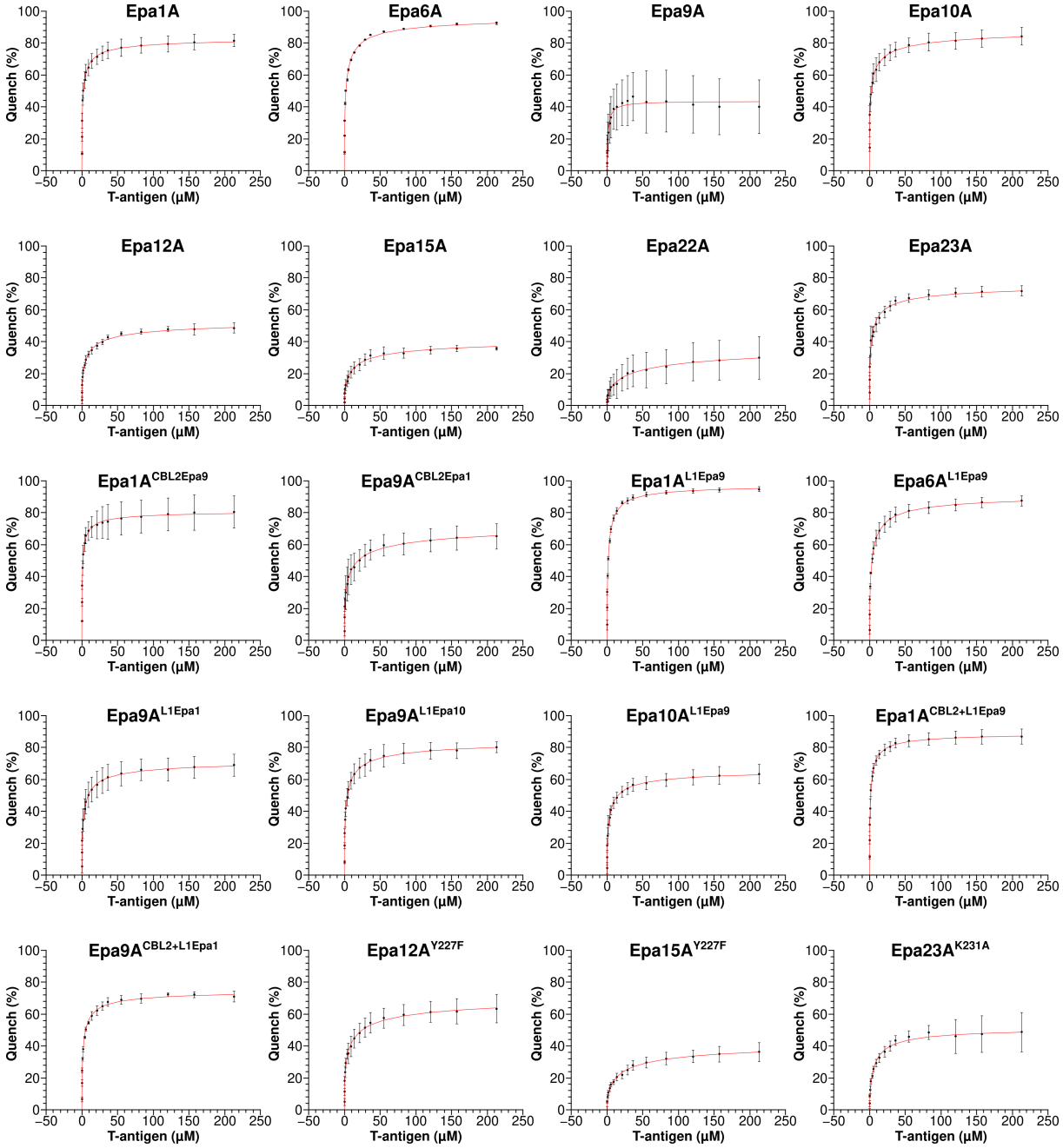

Figure A1: FTS measurements of natural and mutated EpaA domains with T-antigen.
FTS measurements were used to calculate K_D values for the tested EpaA variants. By adding increasing amounts of ligands (T-antigen), the amount of fluorescent light emitted from the protein was reduced. This reduction (quench) was plotted against the ligand concentration to calculate the dissociation constants, using a nonlinear fit.

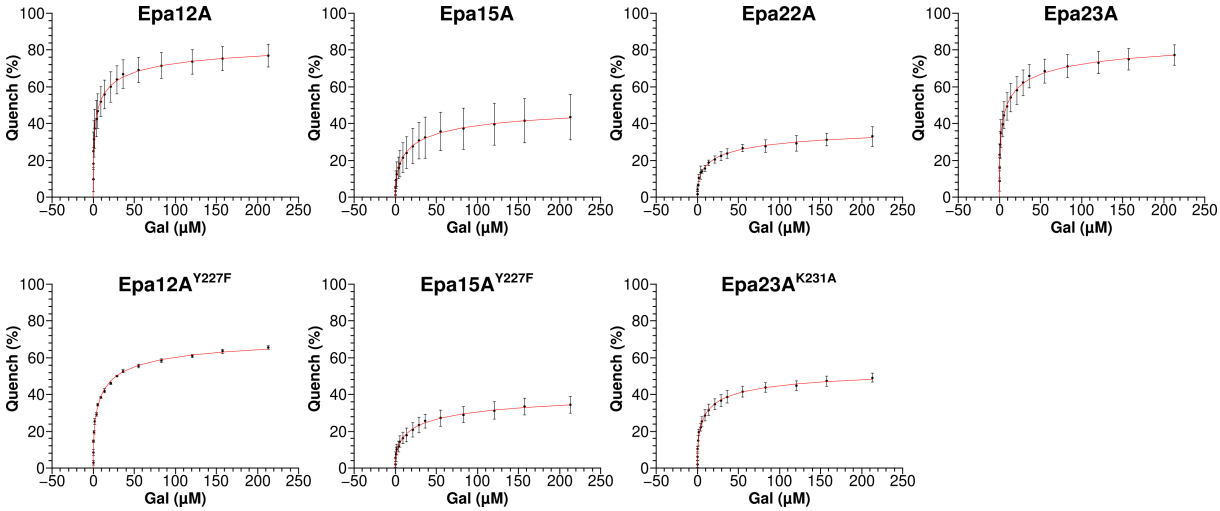

Figure A2: FTS measurements of natural and mutated EpaA domains with galactose.
FTS measurements with galactose were used to calculate K_D values for Epa12A, Epa15A, Epa22A, Epa23A and their respective mutants. The EpaA variants were titrated with galactose and the resulting quench was plotted against the ligand concentration. Dissociation constants were calculated using a nonlinear fit.

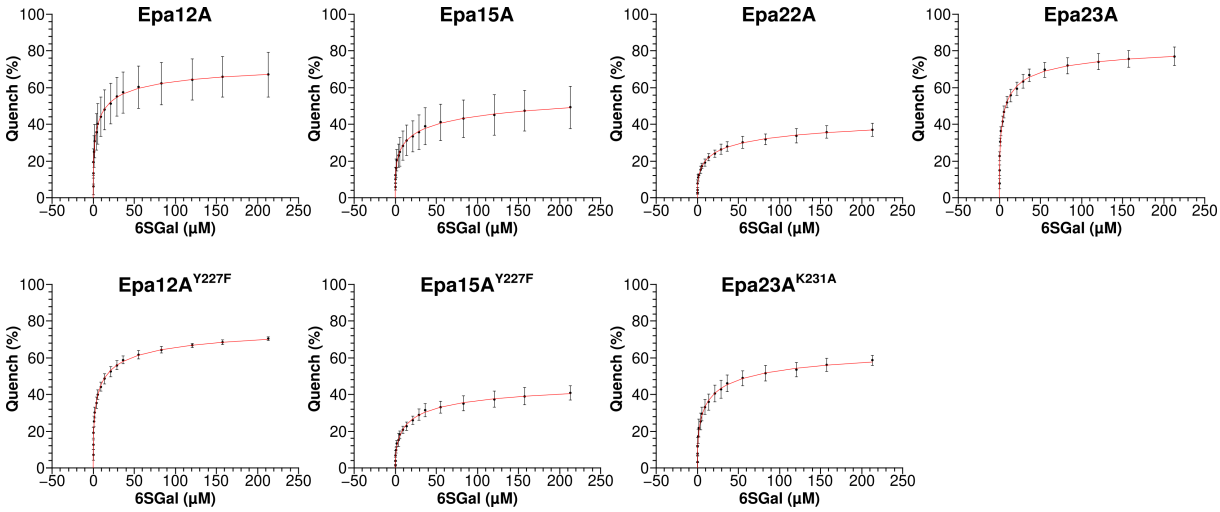

Figure A3: FTS measurements of different natural and mutated EpaA domains with sulfogalactose.
FTS measurements with sulfogalactose were used to calculate K_D values for Epa12A, Epa15A, Epa22A, Epa23A and their respective mutants. The EpaA variants were titrated with sulfogalactose and the resulting quench was plotted against the ligand concentration. Dissociation constants were calculated using a nonlinear fit.

11.2 Overview of *in vivo* and *in vitro* analysis results

Table A1: Results of *in vivo* and *in vitro* analyses of different EpaA variants.

Variant	Host cell adhesion				Glycan array		FTS	Reference
	Caco-2 (%)	TCC-SUP (%)	HeLa (%)	Behavior	Cluster analysis	TDA[2]	K_D (µM)[1]	
control	2	1	11	-	-	-	-	Diderrich et al. 2015
no A	4	1	13	-	-	-	-	Diderrich et al. 2015
Epa1A	59	44	54	natural	natural	natural	1.1 ± 0.1	Diderrich et al. 2015
Epa1A^{D165A}	11[3]	ND[4]	ND	ND	ND	ND	ND	Hoffmann et al. 2020
Epa1A^{C78S}	6[3]	ND	ND	ND	ND	ND	ND	Hoffmann et al. 2020
Epa1A^{W198A}	22[3]	ND	ND	ND	ND	ND	ND	Hoffmann et al. 2020
Epa1A^{W198H}	53[3]	ND	ND	ND	ND	ND	ND	Hoffmann et al. 2020
Epa1A^{W198Y}	67[3]	ND	ND	ND	ND	ND	ND	Hoffmann et al. 2020
Epa1A^{CBL2Epa2}	36[3]	ND	ND	novel	novel	novel	ND	Maestre-Reyna et al. 2012
Epa1A^{CBL2Epa3}	6[3]	ND	ND	donor	novel	novel	30.0 ± 6.1[3]	Maestre-Reyna et al. 2012
Epa1A^{CBL2Epa6}	44[3]	ND	ND	recipient/donor	novel	novel	1.7 ± 0.4[3]	Maestre-Reyna et al. 2012
Epa1A^{CBL2Epa9}	51	43	52	recipient	novel	recipient/donor	0.8 ± 0.1	Hoffmann et al. 2020
Epa1A^{L1Epa9}	67	42	56	recipient	recipient	recipient	1.7 ± 0.1	this work
Epa1A$^{CBL2+L1Epa9}$	7	2	20	novel	ND	ND	1.2 ± 0.1	this work
Epa2A	10[3]	ND	ND	natural	natural	natural	ND	Diderrich et al. 2015
Epa2A^{CBL2Epa1}	6[3]	ND	ND	recipient	ND	ND	ND	Hoffmann et al. 2020

Continuation on next page.

Variant	Host cell adhesion				Glycan array		FTS	Reference
	Caco-2 (%)	TCC-SUP (%)	HeLa (%)	Behavior	Cluster analysis	TDA[2]	K_D (µM)[1]	
Epa2A^{CBL2Epa3}	5[3]	ND	ND	recipient/donor	ND	ND	ND	Hoffmann et al. 2020
Epa3A	9[3]	ND	ND	natural	natural	natural	ND	Diderrich et al. 2015
Epa3A^{CBL2Epa1}	8[3]	ND	ND	recipient	novel	ND	ND	Hoffmann et al. 2020
Epa3A^{CBL2Epa2}	9[3]	ND	ND	recipient/donor	novel	ND	ND	Hoffmann et al. 2020
Epa6A	49	29	41	natural	natural	natural	2.0 ± 0.2	Diderrich et al. 2015
Epa6A^{CBL2Epa1}	28[3]	ND	ND	novel	ND	ND	ND	Hoffmann et al. 2020
Epa6A^{L1Epa9}	13	2	19	donor	recipient	recipient	2.7 ± 0.2	this work
Epa9A	22	4	9	natural	natural	natural	1.4 ± 0.2	Diderrich et al. 2015
Epa9A^{CBL2Epa1}	11	1	9	recipient	donor	recipient	5.1 ± 0.9	Hoffmann et al. 2020
Epa9A^{L1Epa1}	61	41	56	donor	recipient	recipient/novel	2.8 ± 0.3	this work
Epa9A$^{CBL2+L1Epa1}$	7	2	19	novel[5]	ND	ND	1.9 ± 0.2	this work
Epa9A^{L1Epa10}	11	2	18	recipient/donor	recipient	recipient	2.4 ± 0.2	this work
Epa10A	21	3	15	natural	natural	natural	1.6 ± 0.2	Diderrich et al. 2015
Epa10A^{CBL2Epa1}	8	2	14	novel	ND	ND	ND	this work
Epa10A^{L1Epa1}	65	31	50	donor	ND	ND	ND	this work
Epa10A$^{CBL2+L1Epa1}$	6	2	11	novel[5]	ND	ND	ND	this work
Epa10A^{L1Epa9}	12	3	15	recipient/donor	donor	recipient/novel	2.3 ± 0.2	this work
Epa12A	ND	11	ND	natural	natural	natural	5.6 ± 0.7	Diderrich et al. 2015
Epa12A^{Y227F}	ND	8	ND	natural	ND	ND	5.9 ± 1.1	this work
Epa15A	ND	28	ND	natural	natural	natural	9.6 ± 2.1	Diderrich et al. 2015

Continuation on next page.

Variant	Host cell adhesion				Glycan array		FTS	Reference
	Caco-2 (%)	TCC-SUP (%)	HeLa (%)	Behavior	Cluster analysis	TDA[2]	K_D (µM)[1]	
Epa15A^{Y227F}	ND	23	ND	natural	ND	ND	17.7 ± 3.8	this work
Epa22A	ND	9	ND	natural	natural	natural	35.4 ± 13.30	Diderrich et al. 2015
Epa23A	ND	12	ND	natural	natural	natural	2.3 ± 1.0	Diderrich et al. 2015
Epa23A^{K231A}	ND	10	ND	natural	ND	ND	5.8 ± 0.7	this work

[1] Terminal disaccharide analysis

[2] Measured with T-antigen (Galβ1-3GalNAc)

[3] Taken from reference

[4] Not determined

[5] Shows host cell adhesion comparable to its corresponding CBL2 exchange variant.

11.3 Sequences of produced adhesion domains

Epa1A

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH MTSSNDISLA SKDPTTFPLG CSPDITTPKK GLSMELYSYD
        70         80         90        100        110        120
FRKKGSYPCW DAAYLDPNYP RTGYKSHRLL AKVDGVTGNI NFYYHATKGC TPQLGHLPAS
       130        140        150        160        170        180
YNYPKPLTMT NFTMLLYGYF RPKVTGFHTF TISADDLLFV NFGAGNAFDC CRRDSSADHF
       190        200        210        220        230        240
GNYQAYAIWG SKTAKDELTV HLDAGVYYPI RLFYNNREYD GALSFTFKTE SNENTVSDFS
       250        260
EYFFSLDDTE EGCPGLISYD SS
```

Epa6A

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH MKDDYSSSLS NNNLGWTDPT EFPLGCSPNV TTPKNGLSME
        70         80         90        100        110        120
LYSYDYLKSG SNPCWDAAYL DPNYPRTGYK SHRLLAKVEN VAGNINFYYH APMGCTSLFD
       130        140        150        160        170        180
TLPQAYNYRT PLTMTNFTML LYGYFKPKVT GYHTFTISAD DLLFVNFGAG NAFDCCKRES
       130        140        150        160        170        180
SADDFGNYQA YAVWGSQTAK DDLTVHLDAG LYYPIRIFFN NRDNDGALSL TLKTESDPNP
       190        200
VIDFSDYFYS FDDTKDGCPG LVSYDTS
```

Epa9A

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH MASDITPFAH YPRPEGCSSP PNAVSVGLHM DLYNYPYLYV
        70         80         90        100        110        120
KNPRTGFTND TDSDADGETD GDSAGGIEGR AGQCWNPEYQ DPNFPRYGYK KYGSFGSSDH
       130        140        150        160        170        180
VNGKISWDHN EFKEGCKPIM ARLPTAYNYP AKITFSNFTM VLSGYFKPKS TGLYKFEIHA
       130        140        150        160        170        180
DDFILFNFGS KNAFECCNRE ESIDNFGPYV AYAMWPNEAD QELEVYLFED SYYPIRLFYN
       190        200        210        220
NRDYHSKFMV GFYPPNTEEI TYDFDGYLYM LDDTGNECKD SIRYKTVCDD
```

Epa10A

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MDFTPFAHYP  RPEGCSSPPN  SVSVGLHMNL  YNYPYLYVKP
        70          80          90         100         110         120
STKGMFTNDT  NSDSDGETDG  DSAGGIEGRA  GQCWNPEYQD  PNFPRFGYKQ  YGEFGASDNV
       130         140         150         160         170         180
NGDISWDHNE  FTEGCKPVLA  ALPPGYNYPD  EITFSNFTMV  LSGYFKPKTS  GVYKFELKAD
       130         140         150         160         170         180
DFILFNFGAK  NAFECCNREE  SIDNFGPYVA  YAMWPNQADQ  ELEVYLFEDS  YYPLRLFYNN
       190         200         210         220
RDYHSKFKIG  FYPPGQTTIT  FDFDGYLFML  DDTGNECRDS  IRYRTVCDD
```

Epa12A

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MLEPDNSLQI  FDSFLSNPSK  YPLGCSPKIT  NPKKGLSMEL
        70          80          90         100         110         120
YSYPYRKKGS  HPCWDPAYLD  PNFPRVGYKK  NKLIARVDGV  SGDINFNFHP  KRQCTPIADY
       130         140         150         160         170         180
LPPNFNYNEP  ITTTNFTMLL  YGYFKPKVTG  LHTFDISADD  LLFMNFGAGN  AFDCCRRDST
       130         140         150         160         170         180
ADTFGNYVAW  AIWGRRIVRN  KLTVRLDKGI  YYPLRLFFNN  RDYYGQMRLT  FKTEHGSERI
       190
TDFSDYFFSV  DDTSEGCP
```

Epa15A

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MYHDPTTFPL  GCSPEFTKTK  KGLSMELYRY  DYLPPGSYPC
        70          80          90         100         110         120
WDSAYLNPSY  PRTGYKAKKL  IATVDGVSGD  INFKFNPKFG  CKAIPDYLPS  NFNYHEPITI
       130         140         150         160         170         180
TNFTMILYGY  FMPKTTAFHT  FYVTADDLLF  MNFGAGNAFD  CCRREETADK  FGNYAAYSVW
       130         140         150         160         170         180
GKKSLKNELT  VYLHTGVYYP  IRLFYNNRDY  IAQLDIRFKT  EHSNAIITNF  TDYFYSVDDS
       190         200
SLGCPGLITY  EKQCADVQSP
```

Epa23A

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MIPNPFEGTI  NNFPSGCSPN  HGLYNRGLTM  ELYNYSYIHP
        70          80          90         100         110         120
QSIQSYFNTS  TLKKMEKGQC  WDPSYLDVNY  PRTGYKTHNR  FAKVNGIDGI  LDFEFNPTRS
       130         140         150         160         170         180
CVPSKGQLPQ  NYNYPLQFTL  SNFTMLLYGY  FKPKVTAKHT  FTIFADDLLF  LNFGAGNAFD
       130         140         150         160         170         180
CCQQQDTIDD  FGNYQAYALW  GSDTQQNTLT  VNLDANIYYP  IRMFYNNRDF  KGALNMYFTT
       190         200         210
DESNTKINDF  SGYLFNIPDS  SEGCPAHISY  ETECGNVSG
```

Epa1A^{CBL2Epa9}

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MTSSNDISLA  SKDPTTFPLG  CSPDITTPKK  GLSMELYSYD
        70          80          90         100         110         120
FRKKGSYPCW  DAAYLDPNYP  RTGYKSHRLL  AKVDGVTGNI  NFYYHATKGC  TPQLGHLPAS
       130         140         150         160         170         180
YNYPKPLTMT  NFTMLLYGYF  RPKVTGFHTF  TISADDLLFV  NFGAGNAFDC  CRRDSSADHF
       130         140         150         160         170         180
GNYQAYAIWG  SKTAKDELTV  HLDAGVYYPI  RLFYNNRDND  GALSFTFKTE  SNENTVSDFS
       190         200         210         220         230
EYFFSLDDTE  EGCPGLISYD  SSCASVKTSK  IIGIDYHTET  PNENLVPITK  TIYHLG
```

Epa9A^{CBL2Epa1}

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MTSSNDISLA  SKDPTTFPLG  CSPDITTPKK  GLSMELYSYD
        70          80          90         100         110         120
FRKKGSYPCW  DAAYLDPNYP  RTGYKSHRLL  AKVDGVTGNI  NFYYHATKGC  TPQLGHLPAS
       130         140         150         160         170         180
YNYPKPLTMT  NFTMLLYGYF  RPKVTGFHTF  TISADDLLFV  NFGAGNAFDC  CRRDSSADHF
       130         140         150         160         170         180
GNYQAYAIWG  SKTAKDELTV  HLDAGVYYPI  RLFYNNRDYH  GALSFTFKTE  SNENTVSDFS
       190         200
EYFFSLDDTE  EGCPGLISYD  SS
```

Epa1A^{L1Epa9}

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MTSSNDISLA  SKDPTTFPLG  CSPDITTPKK  GLSMELYSYD
        70          80          90         100         110         120
FLYVKNPRTG  FTNDTDSDAD  GETDGDSAGG  IEGRAGQCWD  AAYLDPNYPR  TGYKSHRLLA
       130         140         150         160         170         180
KVDGVTGNIN  FYYHATKGCT  PQLGHLPASY  NYPKPLTMTN  FTMLLYGYFR  PKVTGFHTFT
       130         140         150         160         170         180
ISADDLLFVN  FGAGNAFDCC  RRDSSADHFG  NYQAYAIWGS  KTAKDELTVH  LDAGVYYPIR
       190         200         210         220         230
LFYNNREYDG  ALSFTFKTES  NENTVSDFSE  YFFSLDDTEE  GCPGLISYDS  S
```

Epa6A^{L1Epa9}

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MKDDYSSSLS  NNNLGWTDPT  EFPLGCSPNV  TTPKNGLSME
        70          80          90         100         110         120
LYSYDYLYVK  NPRTGFTNDT  DSDADGETDG  DSAGGIEGRA  GQCWDAAYLD  PNYPRTGYKS
       130         140         150         160         170         180
HRLLAKVENV  AGNINFYYHA  PMGCTSLFDT  LPQAYNYRTP  LTMTNFTMLL  YGYFKPKVTG
       130         140         150         160         170         180
YHTFTISADD  LLFVNFGAGN  AFDCCKRESS  ADDFGNYQAY  AVWGSQTAKD  DLTVHLDAGL
       190         200         210         220         230
YYPIRIFFNN  RDNDGALSLT  LKTESDPNPV  IDFSDYFYSF  DDTKDGCPGL  VSYDTS
```

Epa9A^{L1Epa1}

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MDFTPFAHYP  RPEGCSSPPN  SVSVGLHMNL  YNYPYLYVKN
        70          80          90         100         110         120
PRTGFTNDTD  SDADGETDGD  SAGGIEGRAG  QCWNPEYQDP  NFPRFGYKQY  GEFGASDNVN
       130         140         150         160         170         180
GDISWDHNEF  TEGCKPVLAA  LPPGYNYPDE  ITFSNFTMVL  SGYFKPKTSG  VYKFELKADD
       130         140         150         160         170         180
FILFNFGAKN  AFECCNREES  IDNFGPYVAY  AMWPNQADQE  LEVYLFEDSY  YPLRLFYNNR
       190         200         210         220         230
DYHSKFKIGF  YPPGQTTITF  DFDGYLFMLD  DTGNECRDSI
```

Epa9A^{L1Epa10}

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MASDITPFAH  YPRPEGCSSP  PNAVSVGLHM  DLYNYPYRKK
        70          80          90         100         110         120
GSYPCWNPEY  QDPNFPRYGY  KKYGSFGSSD  HVNGKISWDH  NEFKEGCKPI  MARLPTAYNY
       130         140         150         160         170         180
PAKITFSNFT  MVLSGYFKPK  STGLYKFEIH  ADDFILFNFG  SKNAFECCNR  EESIDNFGPY
       130         140         150         160         170         180
VAYAMWPNEA  DQELEVYLFE  DSYYPIRLFY  NNRDYHSKFM  VGFYPPNTEE  ITYDFDGYLY
       190         200
MLDDTGNECK  DSIRYKTVCD  D
```

Epa10A^{L1Epa9}

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MASDITPFAH  YPRPEGCSSP  PNAVSVGLHM  DLYNYPYLYV
        70          80          90         100         110         120
KPSTKGMFTN  DTNSDSDGET  DGDSAGGIEG  RAGQCWNPEY  QDPNFPRYGY  KKYGSFGSSD
       130         140         150         160         170         180
HVNGKISWDH  NEFKEGCKPI  MARLPTAYNY  PAKITFSNFT  MVLSGYFKPK  STGLYKFEIH
       130         140         150         160         170         180
ADDFILFNFG  SKNAFECCNR  EESIDNFGPY  VAYAMWPNEA  DQELEVYLFE  DSYYPIRLFY
       190         200         210         220         230
NNRDYHSKFM  VGFYPPNTEE  ITYDFDGYLY  MLDDTGNECK  DSIRYKTVCD  D
```

Epa1A$^{CBL2+L1Epa9}$

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MTSSNDISLA  SKDPTTFPLG  CSPDITTPKK  GLSMELYSYD
        70          80          90         100         110         120
FLYVKNPRTG  FTNDTDSDAD  GETDGDSAGG  IEGRAGQCWD  AAYLDPNYPR  TGYKSHRLLA
       130         140         150         160         170         180
KVDGVTGNIN  FYYHATKGCT  PQLGHLPASY  NYPKPLTMTN  FTMLLYGYFR  PKVTGFHTFT
       130         140         150         160         170         180
ISADDLLFVN  FGAGNAFDCC  RRDSSADHFG  NYQAYAIWGS  KTAKDELTVH  LDAGVYYPIR
       190         200         210         220         230
LFYNNRDYHG  ALSFTFKTES  NENTVSDFSE  YFFSLDDTEE  GCPGLISYDS  S
```

Epa9A$^{CBL2+L1Epa1}$

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MASDITPFAH  YPRPEGCSSP  PNAVSVGLHM  DLYNYPYRKK
        70          80          90         100         110         120
GSYPCWNPEY  QDPNFPRYGY  KKYGSFGSSD  HVNGKISWDH  NEFKEGCKPI  MARLPTAYNY
       130         140         150         160         170         180
PAKITFSNFT  MVLSGYFKPK  STGLYKFEIH  ADDFILFNFG  SKNAFECCNR  EESIDNFGPY
       130         140         150         160         170         180
VAYAMWPNEA  DQELEVYLFE  DSYYPIRLFY  NNREYDSKFM  VGFYPPNTEE  ITYDFDGYLY
       190         200
MLDDTGNECK  DSIRYKTVCD  D
```

Epa12A^{Y227F}

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MLEPDNSLQI  FDSFLSNPSK  YPLGCSPKIT  NPKKGLSMEL
        70          80          90         100         110         120
YSYPYRKKGS  HPCWDPAYLD  PNFPRVGYKK  NKLIARVDGV  SGDINFNFHP  KRQCTPIADY
       130         140         150         160         170         180
LPPNFNYNEP  ITTTNFTMLL  YGYFKPKVTG  LHTFDISADD  LLFMNFGAGN  AFDCCRRDST
       130         140         150         160         170         180
ADTFGNYVAW  AIWGRRIVRN  KLTVRLDKGI  YYPLRLFFNN  RDFYGQMRLT  FKTEHGSERI
       190         200         210         220
TDFSDYFFSV  DDTSEGCPGL  ITYESECADV  KSSTVLETDY  ITIQAEKE
```

Epa15A^{Y227F}

```
        10          20          30          40          50          60
MGSSHHHHHH  SSGLVPRGSH  MYGYHDPTTF  PLGCSPEFTK  TKKGLSMELY  RYDYLPPGSY
        70          80          90         100         110         120
PCWDSAYLNP  SYPRTGYKAK  KLIATVDGVS  GDINFKFNPK  FGCKAIPDYL  PSNFNYHEPI
       130         140         150         160         170         180
TITNFTMILY  GYFMPKTTAF  HTFYVTADDL  LFMNFGAGNA  FDCCRREETA  DKFGNYAAYS
       130         140         150         160         170         180
VWGKKSLKNE  LTVYLHTGVY  YPIRLFYNNR  DFIAQLDIRF  KTEHSNAIIT  NFTDYFYSVD
       190         200
DSSLGCPGLI  TYEKQCADVQ  SP
```

Epa23A^{K231A}

10	20	30	40	50	60
MGSSHHHHHH	SSGLVPRGSH	MIPNPFEGTI	NNFPSGCSPN	HGLYNRGLTM	ELYNYSYIHP
70	80	90	100	110	120
QSIQSYFNTS	TLKKMEKGQC	WDPSYLDVNY	PRTGYKTHNR	FAKVNGIDGI	LDFEFNPTRS
130	140	150	160	170	180
CVPSKGQLPQ	NYNYPLQFTL	SNFTMLLYGY	FKPKVTAKHT	FTIFADDLLF	LNFGAGNAFD
130	140	150	160	170	180
CCQQQDTIDD	FGNYQAYALW	GSDTQQNTLT	VNLDANIYYP	IRMFYNNRDF	AGALNMYFTT
190	200	210	220		
DESNTKINDF	SGYLFNIPDS	SEGCPAHISY	ETECGNVSG		

11.4 Glycan arrays

Table A2: Glycan array v5.4

Chart ID	Glycan ID	Glycan structure
1	7Sp8	Gala-Sp8
2	8Sp8	Glca-Sp8
3	9Sp8	Mana-Sp8
4	10Sp8	GalNAca-Sp8
5	10Sp15	GalNAca-Sp15
6	11Sp8	Fuca-Sp8
7	11Sp9	Fuca-Sp9
8	12Sp8	Rhaa-Sp8
9	13Sp8	Neu5Aca-Sp8
10	13Sp11	Neu5Aca-Sp11
11	14Sp8	Neu5Acb-Sp8
12	15Sp8	Galb-Sp8
13	16Sp8	Glcb-Sp8
14	17Sp8	Manb-Sp8
15	18Sp8	GalNAcb-Sp8
16	19Sp0	GlcNAcb-Sp0
17	19Sp8	GlcNAcb-Sp8
18	20Sp8	GlcN[Gc]b-Sp8
19	21Sp8	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAca-Sp8
20	21Sp14	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAca-Sp14
21	22Sp8	GlcNAcb1-6(GlcNAcb1-4)(GlcNAcb1-3)GlcNAc-Sp8
22	23Sp0	[6S][3S]Galb1-4[6S]GlcNAcb-Sp0
23	24Sp0	[6S][3S]Galb1-4GlcNAcb-Sp0
24	25Sp0	[3S]Galb1-4(Fuca1-3)[6S]Glcb-Sp0
25	26Sp8	[3S]Galb1-4Glcb-Sp8
26	27Sp0	[3S]Galb1-4[6S]Glcb-Sp0
27	27Sp8	[3S]Galb1-4[6S]Glcb-Sp8
28	29Sp8	[3S]Galb1-3(Fuca1-4)GlcNAcb-Sp8
29	30Sp8	[3S]Galb1-3GalNAca-Sp8
30	31Sp0	[3S]Galb1-3GlcNAcb-Sp0
31	31Sp8	[3S]Galb1-3GlcNAcb-Sp8
32	32Sp0	[3S]Galb1-4(Fuca1-3)GlcNAcb-Sp0
33	32Sp8	[3S]Galb1-4(Fuca1-3)GlcNAcb-Sp8
34	33Sp0	[3S]Galb1-4[6S]GlcNAcb-Sp0
35	33Sp8	[3S]Galb1-4[6S]GlcNAcb-Sp8
36	34Sp0	[3S]Galb1-4GlcNAcb-Sp0
37	34Sp8	[3S]Galb1-4GlcNAcb-Sp8
38	35Sp8	[3S]Galb-Sp8
39	37Sp0	[6S][4S]Galb1-4GlcNAcb-Sp0
40	38Sp8	[4S]Galb1-4GlcNAcb-Sp8
41	39Sp8	[6P]Mana-Sp8

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
42	40Sp0	[6S]Galb1-4Glcb-Sp0
43	40Sp8	[6S]Galb1-4Glcb-Sp8
44	41Sp8	[6S]Galb1-4GlcNAcb-Sp8
45	42Sp8	[6S]Galb1-4[6S]Glcb-Sp8
46	43Sp8	Neu5Aca2-3[6S]Galb1-4GlcNAcb-Sp8
47	44Sp8	[6S]GlcNAcb-Sp8
48	46Sp8	Neu5,9Aca-Sp8
49	47Sp8	Neu5,9Aca2-6Galb1-4GlcNAcb-Sp8
50	48Sp12	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
51	48Sp13	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13
52	49Sp12	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
53	49Sp13	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13
54	50Sp12	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
55	51Sp12	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
56	51Sp13	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13
57	51Sp24	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24
58	52Sp9	Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9
59	53Sp9	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp9
60	54Sp8	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb-Sp8
61	56Sp8	Fuca1-2Galb1-3GalNAca-Sp8
62	56Sp14	Fuca1-2Galb1-3GalNAca-Sp14
63	57Sp0	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0
64	57Sp9	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp9
65	58Sp8	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp8
66	58Sp10	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10
67	60Sp0	Fuca1-2Galb1-3GlcNAcb-Sp0
68	60Sp8	Fuca1-2Galb1-3GlcNAcb-Sp8
69	61Sp0	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
70	62Sp0	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
71	64Sp0	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0
72	64Sp8	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8
73	65Sp0	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
74	66Sp0	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
75	67Sp0	Fuca1-2Galb1-4GlcNAcb-Sp0
76	67Sp8	Fuca1-2Galb1-4GlcNAcb-Sp8
77	68Sp0	Fuca1-2Galb1-4Glcb-Sp0
78	69Sp8	Fuca1-2Galb-Sp8
79	70Sp8	Fuca1-3GlcNAcb-Sp8
80	72Sp8	Fuca1-4GlcNAcb-Sp8
81	73Sp8	Fucb1-3GlcNAcb-Sp8

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
82	74Sp0	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0
83	75Sp0	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0
84	76Sp0	[3S]Galb1-4(Fuca1-3)Glcb-Sp0
85	77Sp0	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0
86	77Sp8	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8
87	78Sp0	GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0
88	79Sp8	GlcNAcb1-3Galb1-3GalNAca-Sp8
89	80Sp8	GalNAca1-3(Fuca1-2)Galb-Sp8
90	80Sp18	GalNAca1-3(Fuca1-2)Galb-Sp18
91	81Sp8	GalNAca1-3GalNAcb-Sp8
92	82Sp8	GalNAca1-3Galb-Sp8
93	84Sp8	GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8
94	85Sp8	GalNAcb1-3GalNAca-Sp8
95	86Sp8	GalNAcb1-3(Fuca1-2)Galb-Sp8
96	87Sp0	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0
97	90Sp0	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0
98	92Sp0	GalNAcb1-4GlcNAcb-Sp0
99	92Sp8	GalNAcb1-4GlcNAcb-Sp8
100	93Sp8	Gala1-2Galb-Sp8
101	94Sp0	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0
102	94Sp8	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp8
103	95Sp0	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0
104	95Sp8	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp8
105	96Sp0	Gala1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0
106	97Sp0	Gala1-3(Fuca1-2)Galb1-4Glcb-Sp0
107	98Sp8	Gala1-3(Fuca1-2)Galb-Sp8
108	98Sp18	Gala1-3(Fuca1-2)Galb-Sp18
109	99Sp8	Gala1-4(Gala1-3)Galb1-4GlcNAcb-Sp8
110	100Sp8	Gala1-3GalNAca-Sp8
111	100Sp16	Gala1-3GalNAca-Sp16
112	101Sp8	Gala1-3GalNAcb-Sp8
113	102Sp8	Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8
114	103Sp0	Gala1-3Galb1-3GlcNAcb-Sp0
115	104Sp8	Gala1-3Galb1-4GlcNAcb-Sp8
116	105Sp0	Gala1-3Galb1-4Glcb-Sp0
117	105Sp10	Gala1-3Galb1-4Glc-Sp10
118	106Sp8	Gala1-3Galb-Sp8
119	107Sp8	Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8
120	108Sp0	Gala1-4Galb1-4GlcNAcb-Sp0
121	108Sp8	Gala1-4Galb1-4GlcNAcb-Sp8
122	109Sp0	Gala1-4Galb1-4Glcb-Sp0
123	110Sp8	Gala1-4GlcNAcb-Sp8
124	111Sp8	Gala1-6Glcb-Sp8
125	112Sp8	Galb1-2Galb-Sp8
126	113Sp0	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
127	114Sp0	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4GlcNAcb-Sp0
128	115Sp0	Galb1-3(Fuca1-4)GlcNAcb-Sp0
129	115Sp8	Galb1-3(Fuca1-4)GlcNAc-Sp8
130	116Sp8	Fuca1-4(Galb1-3)GlcNAcb-Sp8
131	119Sp8	Galb1-4GlcNAcb1-6GalNAca-Sp8
132	119Sp14	Galb1-4GlcNAcb1-6GalNAca-Sp14
133	120Sp8	GlcNAcb1-6(Galb1-3)GalNAca-Sp8
134	120Sp14	GlcNAcb1-6(Galb1-3)GalNAca-Sp14
135	121Sp8	Neu5Aca2-6(Galb1-3)GalNAca-Sp8
136	121Sp14	Neu5Aca2-6(Galb1-3)GalNAca-Sp14
137	122Sp8	Neu5Acb2-6(Galb1-3)GalNAca-Sp8
138	123Sp10	Neu5Aca2-6(Galb1-3)GlcNAcb1-4Galb1-4Glcb-Sp10
139	126Sp8	Galb1-3GalNAca-Sp8
140	126Sp14	Galb1-3GalNAca-Sp14
141	126Sp16	Galb1-3GalNAca-Sp16
142	127Sp8	Galb1-3GalNAcb-Sp8
143	128Sp0	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0
144	129Sp0	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0
145	130Sp8	Galb1-3GalNAcb1-4Galb1-4Glcb-Sp8
146	131Sp8	Galb1-3Galb-Sp8
147	132Sp0	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0
148	133Sp10	Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10
149	134Sp0	Galb1-3GlcNAcb-Sp0
150	134Sp8	Galb1-3GlcNAcb-Sp8
151	136Sp0	Galb1-4(Fuca1-3)GlcNAcb-Sp0
152	136Sp8	Galb1-4(Fuca1-3)GlcNAcb-Sp8
153	137Sp0	Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb-Sp0
154	138Sp0	Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb-Sp0
155	140Sp0	Galb1-4[6S]Glcb-Sp0
156	140Sp8	Galb1-4[6S]Glcb-Sp8
157	141Sp8	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8
158	142Sp8	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8
159	144Sp8	Galb1-4GlcNAcb1-3GalNAca-Sp8
160	144Sp14	Galb1-4GlcNAcb1-3GalNAca-Sp14
161	145Sp0	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
162	146Sp0	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
163	147Sp0	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
164	148Sp0	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0
165	148Sp8	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp8
166	152Sp8	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp8
167	152Sp14	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14
168	153Sp0	Galb1-4GlcNAcb-Sp0
169	153Sp8	Galb1-4GlcNAcb-Sp8

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
170	153Sp23	Galb1-4GlcNAcb-Sp23
171	154Sp0	Galb1-4Glcb-Sp0
172	154Sp8	Galb1-4Glcb-Sp8
173	155Sp8	GlcNAca1-3Galb1-4GlcNAcb-Sp8
174	156Sp8	GlcNAca1-6Galb1-4GlcNAcb-Sp8
175	158Sp8	GlcNAcb1-2Galb1-3GalNAca-Sp8
176	159Sp8	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp8
177	159Sp14	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp14
178	161Sp8	GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAcb-Sp8
179	162Sp8	GlcNAcb1-3GalNAca-Sp8
180	162Sp14	GlcNAcb1-3GalNAca-Sp14
181	163Sp8	GlcNAcb1-3Galb-Sp8
182	164Sp0	GlcNAcb1-3Galb1-4GlcNAcb-Sp0
183	164Sp8	GlcNAcb1-3Galb1-4GlcNAcb-Sp8
184	165Sp0	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
185	166Sp0	GlcNAcb1-3Galb1-4Glcb-Sp0
186	167MDPLys	GlcNAcb1-4-MDPLys
187	168Sp8	GlcNAcb1-6(GlcNAcb1-4)GalNAca-Sp8
188	169Sp8	GlcNAcb1-4Galb1-4GlcNAcb-Sp8
189	170Sp8	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8
190	171Sp8	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8
191	173Sp8	GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8
192	176Sp8	GlcNAcb1-6GalNAca-Sp8
193	176Sp14	GlcNAcb1-6GalNAca-Sp14
194	177Sp8	GlcNAcb1-6Galb1-4GlcNAcb-Sp8
195	178Sp8	Glca1-4Glcb-Sp8
196	179Sp8	Glca1-4Glca-Sp8
197	180Sp8	Glca1-6Glca1-6Glcb-Sp8
198	181Sp8	Glcb1-4Glcb-Sp8
199	182Sp8	Glcb1-6Glcb-Sp8
200	183Sp8	Glc-ol-Sp8
201	184Sp8	GlcAa-Sp8
202	185Sp8	GlcAb-Sp8
203	186Sp8	GlcAb1-3Galb-Sp8
204	187Sp8	GlcAb1-6Galb-Sp8
205	188Sp0	KDNa2-3Galb1-3GlcNAcb-Sp0
206	189Sp0	KDNa2-3Galb1-4GlcNAcb-Sp0
207	190Sp9	Mana1-2Mana1-2Mana1-3Mana-Sp9
208	191Sp9	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9
209	192Sp9	Mana1-2Mana1-3Mana-Sp9
210	198Sp12	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
211	199Sp9	Mana1-6(Mana1-3)Mana-Sp9
212	200Sp9	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
213	203Sp12	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
214	205Sp12	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
215	208Sp0	Manb1-4GlcNAcb-Sp0
216	210Sp0	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
217	211Sp8	[3S]Galb1-4(Fuca1-3)[6S]GlcNAcb-Sp8
218	212Sp0	Fuca1-2[6S]Galb1-4GlcNAcb-Sp0
219	213Sp8	Fuca1-2Galb1-4[6S]GlcNAcb-Sp8
220	216Sp0	Fuca1-2[6S]Galb1-4[6S]Glcb-Sp0
221	220Sp8	Neu5Aca2-3Galb1-3GalNAca-Sp8
222	220Sp14	Neu5Aca2-3Galb1-3GalNAca-Sp14
223	221Sp0	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0
224	223Sp0	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0
225	224Sp0	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0
226	225Sp0	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0
227	227Sp8	Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8
228	230Sp0	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp0
229	230Sp8	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp8
230	231Sp0	GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0
231	232Sp0	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0
232	233Sp8	Neu5Aca2-6(Neu5Aca2-3)GalNAca-Sp8
233	234Sp8	Neu5Aca2-3GalNAca-Sp8
234	235Sp0	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0
235	236Sp8	Neu5Aca2-3Galb1-3[6S]GlcNAc-Sp8
236	237Sp8	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8
237	238Sp0	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
238	240Sp8	Neu5Aca2-3Galb1-4(Neu5Aca2-3Galb1-3)GlcNAcb-Sp8
239	243Sp8	Neu5Aca2-3Galb1-3[6S]GalNAca-Sp8
240	244Sp8	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp8
241	244Sp14	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14
242	245Sp8	Neu5Aca2-3Galb-Sp8
243	246Sp0	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0
244	247Sp0	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0
245	248Sp0	Fuca1-2[6S]Galb1-4Glcb-Sp0
246	249Sp0	Neu5Aca2-3Galb1-3GlcNAcb-Sp0
247	250Sp8	Neu5Aca2-3Galb1-4[6S]GlcNAcb-Sp8
248	251Sp8	Neu5Aca2-3Galb1-4(Fuca1-3)[6S]GlcNAcb-Sp8
249	252Sp0	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
250	253Sp0	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
251	253Sp8	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8
252	254Sp8	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8
253	255Sp8	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8
254	257Sp0	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
255	258Sp0	Neu5Aca2-3Galb1-4GlcNAcb-Sp0

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
256	258Sp8	Neu5Aca2-3Galb1-4GlcNAcb-Sp8
257	259Sp0	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
258	260Sp0	Fuca1-2Galb1-4[6S]Glcb-Sp0
259	261Sp0	Neu5Aca2-3Galb1-4Glcb-Sp0
260	261Sp8	Neu5Aca2-3Galb1-4Glcb-Sp8
261	263Sp8	Neu5Aca2-6GalNAca-Sp8
262	264Sp0	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0
263	267Sp8	Neu5Aca2-6Galb1-4[6S]GlcNAcb-Sp8
264	268Sp0	Neu5Aca2-6Galb1-4GlcNAcb-Sp0
265	268Sp8	Neu5Aca2-6Galb1-4GlcNAcb-Sp8
266	269Sp0	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
267	270Sp0	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
268	271Sp0	Neu5Aca2-6Galb1-4Glcb-Sp0
269	271Sp8	Neu5Aca2-6Galb1-4Glcb-Sp8
270	272Sp8	Neu5Aca2-6Galb-Sp8
271	273Sp8	Neu5Aca2-8Neu5Aca-Sp8
272	274Sp0	Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0
273	275Sp0	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
274	277Sp8	Neu5Acb2-6GalNAca-Sp8
275	278Sp8	Neu5Acb2-6Galb1-4GlcNAcb-Sp8
276	280Sp0	Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
277	281Sp0	Neu5Gca2-3Galb1-3GlcNAcb-Sp0
278	282Sp0	Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
279	283Sp0	Neu5Gca2-3Galb1-4GlcNAcb-Sp0
280	284Sp0	Neu5Gca2-3Galb1-4Glcb-Sp0
281	285Sp0	Neu5Gca2-6GalNAca-Sp0
282	286Sp0	Neu5Gca2-6Galb1-4GlcNAcb-Sp0
283	287Sp8	Neu5Gca-Sp8
284	295Sp14	Neu5Aca2-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14
285	297Sp0	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0
286	298Sp0	Galb1-4(Fuca1-3)[6S]GlcNAcb-Sp0
287	299Sp0	Galb1-4(Fuca1-3)[6S]Glcb-Sp0
288	300Sp0	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
289	301Sp0	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0
290	302Sp0	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0
291	303Sp0	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0
292	305Sp0	[4S][3S]Galb1-4GlcNAcb-Sp0
293	306Sp0	[6S]Galb1-4[6S]GlcNAcb-Sp0
294	307Sp10	[6P]Glcb-Sp10
295	311Sp14	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3)GalNAca-Sp14
296	312Sp8	Galb1-3Galb1-4GlcNAcb-Sp8
297	313Sp12	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
298	314Sp0	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAcb-Sp0

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
299	315Sp0	GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAcb-Sp0
300	316Sp0	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0
301	317Sp0	Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0
302	320Sp8	GalNAcb1-3Galb-Sp8
303	321Sp8	GlcAb1-3GlcNAcb-Sp8
304	322Sp12	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
305	324Sp10	GlcNAcb1-3Man-Sp10
306	325Sp10	GlcNAcb1-4GlcNAcb-Sp10
307	325Sp12	GlcNAcb1-4GlcNAcb-Sp12
308	327Sp10	MurNAcb1-4GlcNAcb-Sp10
309	329Sp10	Mana1-6Manb-Sp10
310	330Sp10	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10
311	331Sp9	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9
312	332Sp9	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9
313	333Sp14	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14
314	336Sp12	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
315	337Sp12	Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
316	340Sp17	Neu5Aca2-8Neu5Acb-Sp17
317	341Sp8	Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8
318	342Sp8	Neu5Gcb2-6Galb1-4GlcNAc-Sp8
319	343Sp19	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
320	345Sp12	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
321	346Sp12	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
322	347Sp20	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
323	351Sp0	Neu5,9Aca2-3Galb1-3GlcNAcb-Sp0
324	352Sp0	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0
325	353Sp0	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
326	354Sp0	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
327	355Sp0	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0
328	356Sp0	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0
329	357Sp0	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
330	358Sp0	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
331	359Sp14	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14
332	360Sp0	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
333	361Sp0	GlcNAca1-4Galb1-4GlcNAcb-Sp0
334	362Sp0	GlcNAca1-4Galb1-3GlcNAcb-Sp0

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
335	363Sp0	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0
336	364Sp0	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
337	365Sp0	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
338	366Sp14	GlcNAca1-4Galb1-3GalNAca-Sp14
339	367Sp12	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
340	368Sp12	Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
341	369Sp12	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAcb-Sp12
342	370Sp12	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAcb-Sp12
343	371Sp12	Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAcb-Sp12
344	372Sp12	Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAcb-Sp12
345	373Sp12	Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
346	374Sp22	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
347	375Sp22	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
348	376Sp22	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
349	379Sp0	[6S]GlcNAcb1-3Galb1-4GlcNAcb-Sp0
350	380Sp0	KDNa2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
351	381Sp0	KDNa2-6Galb1-4GlcNAcb-Sp0
352	382Sp0	KDNa2-3Galb1-4Glcb-Sp0
353	383Sp14	KDNa2-3Galb1-3GalNAca-Sp14
354	384Sp20	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
355	385Sp20	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
356	386Sp20	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
357	387Sp20	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
358	388Sp12	Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
359	389Sp22	Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-6(Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
360	390Sp21	Neu5Aca2-6GlcNAcb1-4GlcNAcb-Sp21
361	391Sp21	Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp21
362	392Sp21	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)Galb1-4Glcb-Sp21
363	393Sp21	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
364	394Sp20	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
365	395Sp20	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
366	396Sp20	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
367	397Sp20	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
368	398Sp20	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
369	399Sp19	Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
370	400Sp14	Neu5Aca2-3Galb1-4GlcNAcb1-3GalNAca-Sp14
371	401Sp14	Neu5Aca2-6Galb1-4GlcNAcb1-3GalNAca-Sp14
372	402Sp14	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14
373	403Sp12	GalNAcb1-4GlcNAcb1-2Mana1-6(GalNAcb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
374	404Sp14	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0
375	405Sp14	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0
376	406Sp0	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp0
377	407Sp21	Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21
378	408Sp21	Galb1-4GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glcb-Sp21
379	409Sp21	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glcb-Sp21
380	410Sp21	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21
381	411Sp21	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
382	412Sp21	GlcNAcb1-2Mana1-6(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
383	413Sp0	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0
384	414Sp0	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0
385	415Sp14	Galb1-3GlcNAcb1-3GalNAca-Sp14
386	416Sp14	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb1-3GalNAca-Sp14
387	417Sp0	GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0
388	418Sp19	Gala1-3Galb1-3GlcNAcb1-2Mana1-6(Gala1-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
389	419Sp19	Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
390	421Sp12	GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
391	422Sp12	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
392	423Sp14	Neu5Aca2-3Galb1-3GlcNAcb1-3GalNAca-Sp14
393	424Sp14	Fuca1-2Galb1-4GlcNAcb1-3GalNAca-Sp14
394	425Sp14	Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14
395	426Sp0	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0
396	427Sp19	Gala1-4Galb1-3GlcNAcb1-2Mana1-6(Gala1-4Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
397	428Sp24	Gala1-4Galb1-4GlcNAcb1-2Mana1-6(Gala1-4Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24
398	429Sp14	Gala1-3Galb1-4GlcNAcb1-3GalNAca-Sp14
399	430Sp0	Galb1-3GlcNAcb1-6Galb1-4GlcNAcb-Sp0
400	431Sp0	Galb1-3GlcNAca1-6Galb1-4GlcNAcb-Sp0
401	432Sp8	GalNAcb1-3Gala1-6Galb1-4Glcb-Sp8
402	434Sp21	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)Glcb-Sp21
403	435Sp21	Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21
404	436Sp0	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0
405	437Sp0	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0
406	438Sp14	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14
407	439Sp14	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14
408	440Sp0	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0
409	441Sp14	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14
410	442Sp14	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14
411	443Sp14	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14
412	444Sp22	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
413	445Sp22	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
414	446Sp19	GlcNAcb1-2(GlcNAcb1-6)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
415	447Sp14	Fuca1-2Galb1-3GlcNAcb1-3GalNAca-Sp14
416	448Sp14	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAca-Sp14
417	449Sp14	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAca-Sp14
418	450Sp14	Gala1-3Galb1-3GlcNAcb1-3GalNAca-Sp14
419	451Sp22	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
420	452Sp22	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
421	453Sp19	Galb1-3GlcNAcb1-6(Galb1-3GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
422	454Sp21	Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21
423	455Sp21	Fuca1-3GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4Glcb-Sp21
424	456Sp21	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
425	457Sp21	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
426	458Sp21	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
427	459Sp21	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
428	460Sp21	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
429	461Sp21	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
430	462Sp21	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
431	463Sp21	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
432	465Sp10	Galb1-4Galb-Sp10
433	466Sp10	Galb1-6Galb-Sp10
434	467Sp8	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb-Sp8
435	468Sp8	GalNAcb1-6GalNAcb-Sp8
436	469Sp0	[6S]Galb1-3GlcNAcb-Sp0
437	470Sp0	[6S]Galb1-3[6S]GlcNAcb-Sp0
438	471Sp12	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
439	472Sp12	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
440	475Sp14	Galb1-4(Fuca1-3)GlcNAcb1-6GalNAca-Sp14
441	476Sp0	Galb1-4GlcNAcb1-2Mana-Sp0
442	478Sp14	Fuca1-2Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)GalNAca-Sp14
443	479Sp14	Gala1-3Fuca1-2Galb1-4GlcNAcb1-3(Gala1-3Fuca1-2Galb1-4GlcNAcb1-6)GalNAc-Sp14
444	480Sp14	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAca-Sp14
445	481Sp0	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0
446	482Sp0	GalNAcb1-4Galb1-4Glcb-Sp0
447	483Sp22	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
448	484Sp22	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
449	485Sp21	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21
450	486Sp22	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
451	487Sp19	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
452	489Sp21	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
453	490Sp21	Neu5Aca2-3Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
454	491Sp21	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
455	492Sp21	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
456	493Sp21	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
457	494Sp21	Neu5Aca2-6Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
458	495Sp21	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
459	496Sp21	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
460	499Sp8	Gala1-3(Fuca1-2)Galb1-3GalNAca-Sp8
461	500Sp8	Gala1-3(Fuca1-2)Galb1-3GalNAcb-Sp8
462	501Sp10	Glca1-6Glca1-6Glca1-6Glcb-Sp10
463	502Sp10	Glca1-4Glca1-4Glca1-4Glcb-Sp10
464	503Sp14	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3)GalNAca-Sp14
465	504Sp24	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
466	505Sp19	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19
467	507Sp24	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
468	508Sp21	Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
469	509Sp21	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21
470	510Sp0	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana-Sp0
471	511Sp14	Neu5Aca2-3Galb1-4GlcNAcb1-6GalNAca-Sp14
472	512Sp14	Neu5Aca2-6Galb1-4GlcNAcb1-6GalNAca-Sp14
473	513Sp14	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3)GalNAca-Sp14
474	514Sp24	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
475	515Sp24	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
476	516Sp19	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19
477	517Sp24	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
478	518Sp21	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
479	519Sp21	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3)Galb1-4Glcb-Sp21
480	520Sp14	Galb1-3GlcNAcb1-6GalNAca-Sp14
481	521Sp14	Gala1-3Galb1-3GlcNAcb1-6GalNAca-Sp14
482	522Sp14	Galb1-3(Fuca1-4)GlcNAcb1-6GalNAca-Sp14
483	523Sp14	Neu5Aca2-3Galb1-3GlcNAcb1-6GalNAca-Sp14
484	524Sp0	[3S]Galb1-3(Fuca1-4)GlcNAcb-Sp0
485	525Sp21	Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-6(Neu5Aca2-3Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21
486	526Sp14	Fuca1-2Galb1-4GlcNAcb1-6GalNAca-Sp14
487	527Sp14	Gala1-3Galb1-4GlcNAcb1-6GalNAca-Sp14
488	528Sp0	Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0
489	529Sp0	Fuca1-2[6S]Galb1-3GlcNAcb-Sp0
490	530Sp14	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAca-Sp14
491	531Sp0	Fuca1-2Galb1-4GlcNAcb1-2Mana-Sp0
492	532Sp0	Fuca1-2Galb1-3[6S]GlcNAcb-Sp0
493	533Sp0	Fuca1-2[6S]Galb1-3[6S]GlcNAcb-Sp0
494	534Sp8	Neu5Aca2-6GalNAcb1-4[6S]GlcNAcb-Sp8
495	535Sp8	GalNAcb1-4(Fuca1-3)[6S]GlcNAcb-Sp8
496	536Sp8	[3S]GalNAcb1-4(Fuca1-3)GlcNAcb-Sp8
497	537Sp14	Fuca1-2Galb1-3GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)GalNAca-Sp14
498	538Sp14	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAca-Sp14
499	539Sp21	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp21
500	540Sp21	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp21
501	541Sp8	Galb1-3GlcNAca1-3Galb1-4GlcNAcb-Sp8
502	542Sp8	Galb1-3[6S]GlcNAcb-Sp8
503	543Sp8	[6S][4S]GalNAcb1-4GlcNAc-Sp8
504	544Sp8	[6S]GalNAcb1-4GlcNAc-Sp8
505	545Sp8	[3S]GalNAcb1-4[3S]GlcNAc-Sp8
506	546Sp8	GalNAcb1-4[6S]GlcNAc-Sp8
507	547Sp8	[3S]GalNAcb1-4GlcNAc-Sp8
508	548Sp10	[4S]GalNAcb-Sp10
509	549Sp0	Galb1-4[6P]GlcNAcb-Sp0
510	550Sp0	[6P]Galb1-4GlcNAcb-SP0
511	551Sp14	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAca-Sp14
512	552Sp0	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana-Sp0
513	553Sp0	Gala1-3Galb1-4GlcNAcb1-2Mana-Sp0
514	554Sp0	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana-Sp0
515	555Sp0	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana-Sp0
516	556Sp0	Galb1-3GlcNAcb1-2Mana-Sp0

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
517	557Sp14	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAca-Sp14
518	558Sp0	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana-Sp0
519	559Sp0	Gala1-3Galb1-3GlcNAcb1-2Mana-Sp0
520	560Sp0	GalNAcb1-4GlcNAcb1-2Mana-Sp0
521	561Sp0	Author: GalNAc anstatt GlcNAc Neu5Aca2-3Galb1-3GlcNAcb1-4Galb1-4Glcb-Sp0
522	563Sp21	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp21
523	564Sp21	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp21
524	565Sp21	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp21
525	566Sp0	Fuca1-4(Galb1-3)GlcNAcb1-2Mana-Sp0
526	567Sp0	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0
527	568Sp0	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAcb-Sp0
528	569Sp21	GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp21
529	570Sp21	Gala1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp21
530	571Sp21	Galb1-3GalNAcb1-3Galb-Sp21
531	572Sp12	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
532	572Sp25	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25
533	573Sp12	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
534	575Sp24	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24
535	576Sp12	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
536	576Sp25	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25
537	577Sp12	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
538	580Sp25	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25
539	581Sp0	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAcb-Sp0
540	582Sp0	Neu5Aca2-8Neu5Gca2-3Galb1-4GlcNAcb-Sp0
541	583Sp0	Neu5Gca2-8Neu5Aca2-3Galb1-4GlcNAcb-Sp0
542	584Sp0	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
543	585Sp0	Neu5Gca2-8Neu5Gca2-6Galb1-4GlcNAcb-Sp0
544	586Sp0	Neu5Aca2-8Neu5Aca2-3Galb1-4GlcNAcb-Sp0
545	587Sp24	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
546	588Sp24	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Mana1-4GlcNAcb1-4GlcNAcb-Sp24
547	589Sp24	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24
548	590Sp14	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-3)GalNAca-Sp14
549	591Sp0	GalNAcb1-3GlcNAcb-Sp0
550	592Sp0	GalNAcb1-4GlcNAcb1-3GalNAcb1-4GlcNAcb-Sp0
551	593Sp25	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25
552	594Sp25	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25
553	595Sp14	GlcNAcb1-3Galb1-3GalNAca-Sp14
554	596Sp14	Galb1-3GlcNAcb1-6(Galb1-3)GalNAca-Sp14
555	598Sp0	[3S]GlcAb1-3Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0
556	599Sp0	[3S]GlcAb1-3Galb1-4GlcNAcb1-2Mana-Sp0
557	600Sp24	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
558	601Sp24	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
559	602Sp21	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp21
560	604Sp24	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
561	605Sp24	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
562	611Sp24	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
563	613Sp24	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24
564	618Sp14	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14
565	619Sp14	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14
566	620Sp14	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14
567	621Sp14	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14

Continuation on next page.

Chart ID	Glycan ID	Glycan structure
568	622Sp14	GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14
569	623Sp14	GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14
570	624Sp14	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14
571	625Sp14	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14
572	626Sp14	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14
573	627Sp14	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14
574	628Sp14	Galb1-4GlcNAcb1-3Galb1-3GalNAca-Sp14
575	630Sp14	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14
576	631Sp14	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14
577	632Sp14	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14
578	633Sp12	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
579	634Sp14	GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14
580	635Sp14	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14
581	636Sp12	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
582	637Sp12	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
583	638Sp12	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
584	639Sp21	GlcNAcb1-3Fuca-Sp21
585	640Sp21	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp21

Table A3: Glycan array with novel glycan structures (AG Unverzagt, University of Bayreuth)

Chart ID	Glycan structure
1	GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
2	GlcNAc-b1,2-(GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
3	GlcNAc-b1,2-Man-a1,3-[GlcNAc-b1,2-(GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
4	GlcNAc-b1,2-(GlcNAc-b1,4)-Man-a1,3-[GlcNAc-b1,2-(GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
5	GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
6	GlcNAc-b1,2-(GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
7	GlcNAc-b1,2-Man-a1,3-[GlcNAc-b1,2-(GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
8	GlcNAc-b1,2-(GlcNAc-b1,4)-Man-a1,3-[GlcNAc-b1,2-(GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
9	GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-(GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
10	GlcNAc-b1,2-(GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-(GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
11	GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-[GlcNAc-b1,2-(GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
12	GlcNAc-b1,2-(GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-[GlcNAc-b1,2-(GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
13	GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-(GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
14	GlcNAc-b1,2-(GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-(GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
15	GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-[GlcNAc-b1,2-(GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
16	GlcNAc-b1,2-(GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-[GlcNAc-b1,2-(GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
17	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
18	Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
19	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-[Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
20	Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-[Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
21	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
22	Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp

Continuation on next page.

Chart ID	Glycan structure
23	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-[Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
24	Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-[Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
25	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
26	Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
27	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-[Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
28	Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(Gal-b1,4-GlcNAc-b1,4)-[GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
29	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
30	Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
31	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-[Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
32	Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-[Gal-b1,4-GlcNAc-b1,2-(Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
33	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
34	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
35	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-[Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
36	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-[Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
37	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
38	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
39	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-[Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
40	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-[Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
41	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
42	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
43	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-[Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp

Continuation on next page.

Chart ID	Glycan structure
44	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-[Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
45	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
46	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
47	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-[Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
48	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-[Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
49	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
50	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
51	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-[Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
52	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-[Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
53	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
54	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
55	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-[Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
56	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-[Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
57	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
58	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
59	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-[Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
60	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-[Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
61	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp

Continuation on next page.

Chart ID	Glycan structure
62	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
63	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,4)-[Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
64	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,4)-Man-a1,3-(GlcNAc-b1,4)-[Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,6)-Man-a1,6]-Man-b1,4-GlcNAc-b1,4-(Fuc-a1,6)-GlcNAc-Sp
65	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
66	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
67	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
68	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
69	Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
70	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
71	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Neu5Ac-a2,3-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
72	Neu5Ac-a2,3-Gal-b1,4-GlcNAc-Sp
73	Neu5Ac-a2,6-Gal-b1,4-GlcNAc-Sp
74	Gal-b1,3-GlcNAc-b1,4-Gal-b1,4-GlcNAc-Sp
75	Fuc-a1,2-Gal-b1,3-GlcNAc-Sp
76	GlcNAc-b1,2-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(GlcNAc-b1,2-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
77	Gal-b1,4-GlcNAc-b1,2-Gal-b1,4-GlcNAc-b1,2-Man-a1,3-(Gal-b1,4-GlcNAc-b1,2-Gal-b1,4-GlcNAc-b1,2-Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
78	Man-a1,2-Man-a1,2-Man-a1,3-(Man-a1,6)-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
79	Man-a1,3-[Man-a1,3-(Man-a1,6)]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp
80	Man-a1,2-Man-a1,3-[Man-a1,2-Man-a1,3-(Man-a1,2-Man-a1,6)]-Man-b1,4-GlcNAc-b1,4-GlcNAc-Sp

Lebenslauf

Diese Seite enthält persönliche Daten und ist daher nicht Teil der Online-Veröffentlichung.

Ein großes Dankeschön...

... Prof. Dr. Hans-Ulrich Mösch für die Bereitstellung und Betreuung meines Projektes und für die Unterstützung dabei, Ordnung in diesen großen Berg aus Daten zu bringen.

... Prof. Dr. Lars-Oliver Essen für die Übernahme der Zweitkorrektur, die vielen hilfreichen Anregungen und die Beschaffung der komplexen Glykane.

... Viktoria Reithofer und den übrigen Mitgliedern der AG Essen für die vielen Tipps zum Thema Kristallisation und vor allem für die Lösung der Proteinstrukturen.

... der AG Borchers, allen voran Hanna, für die Anleitung zum Thema Zellkultur und die Erlaubnis, euer Labor nutzen zu können. Ohne euch wären die *in vivo* Experimente nicht möglich gewesen!

... Ralf Pöschke für das Pipettieren der vielen Kristallisationsansätze und für die unkomplizierte Zusammenarbeit.

... meinen Bachelor- und Masterstudentinnen Larissa, Sabse und Carmen. Ihr habt mir die Betreuung leicht gemacht und habt tolle Arbeit geleistet!

... allen derzeitigen und ehemaligen Mitgliedern der AGs Mösch und Taxis (und Jochen!). Es hat Spaß gemacht mit euch zu arbeiten. Danke auch für die vielen Spieleabende, das Herumkaspern in der Kaffeeküche und natürlich auch die Besuche auf dem Weihnachtsmarkt und beim Stadtfest. Ihr seid die Besten!

... an alle meine Freunde, die das Studium in Marburg wirklich zu einer schönen Zeit gemacht haben. Vor allem meinen Brettspiel-Homies und den Rollenspielern, die mein legendäres Würfelpech ertragen müssen und trotzdem unerschrocken in den Kampf gegen Grolme, Assassinen, Nekromanten, Vampire, Klingonen, Romulaner und andere Gegner ziehen.

... speziell an Didi, die es in ihrer unnachahmlichen Art schafft, Geschichten und Figuren Leben einzuhauchen und immer einen (un)passenden Spruch parat hat. Egal ob Bullshit-Bingo, das neuste Kickstarter-Brettspiel oder ein Gespräch unter Trekkies, du bist immer mit voller Begeisterung dabei. Wie Captain Martok sagen würde: “May’Daq jaHDI’ SuvwI’ juppu’Daj lonbe’!”

... meiner Familie, die immer für mich da ist, mich bedingungslos unterstützt und mich immer dazu motiviert den nächsten Schritt zu machen. Danke für alles!

... an Teresa, die mir mit ihrer Liebe Halt gibt und mich auf den Boden der Tatsachen zurückholt, wenn ich diesen mal verlassen sollte.